# DESIGNING THE LEARNING ANALYTICS CAPABILITY MODEL


JUSTIAN KNOBBOUT

# Designing the Learning Analytics Capability Model

Justian Knobbout

The research presented in this thesis was conducted with the support of a Ph.D. voucher from HU University of Applied Sciences Utrecht, the Netherlands.



ISBN: 978-94-6416-856-3
Cover design and layout: Birgit Vredenburg, Persoonlijk Proefschrift
Printing: Ridderprint, www.ridderprint.nl

# Designing the Learning Analytics Capability Model

**Proefschrift**

ter verkrijging van de graad van doctor

aan de Open Universiteit

op gezag van de rector magnificus

prof. dr. Th. J. Bastiaens

ten overstaan van een door het

College voor promoties ingestelde commissie

in het openbaar te verdedigen

op vrijdag 29 oktober 2021 te Heerlen

om 13:30 uur precies

door

**Justus Hendrik Knobbout**

geboren op 24 januari 1989 te Culemborg

**Promotor**

| | |
|---|---|
| Prof. dr. ir. J.M. Versendaal | Open Universiteit |

**Copromotores**

| | |
|---|---|
| Dr. E.J. van der Stappen | Avans Hogeschool |
| Dr. R. van de Wetering | Open Universiteit |

**Leden beoordelingscommissie**

| | |
|---|---|
| Prof. dr. S. Brinkkemper | Universiteit Utrecht |
| Prof. dr. M. Scheffel | Ruhr-Universität Bochum |
| Prof. dr. M.M. Specht | TU Delft |
| Prof. dr. ir. R.W. Helms | Open Universiteit |
| Prof. dr. ir. C.J. Kreijns | Open Universiteit |

*Difficulties mastered are opportunities won.*

Sir Winston Churchill

## Acknowledgments

Doing Ph.D. research is like riding the Gartner Hype Cycle. It is triggered by reading an interesting paper, in my case about using student data for all kinds of fancy and complex-looking analyses. Having some initial work completed and the results published, nothing seems to go wrong anymore - expectation at its peak. Then reality hits: collecting data takes twice (or three times?) as long as expected, every sentence in written papers comes back with comments and questions, other researchers publish work that is quite similar and - at least in my opinion - much better than mine, and do not get me started on the endless stream of rejected papers. `Disillusionment' is a neat, polite way of describing the mixture of feelings and emotions that popped up from time to time. However, in the final year, there's hope on the horizon. The situation has normalized, and the plateau of productivity is reached - at last. Collecting and structuring all work that has been done over the years is not that much of a hassle anymore and somehow, finishing this Ph.D. thesis produces energy rather than draining it.

Winter 2013. I had a coffee with my former honors program supervisor Johan Versendaal. I finished my master's in Supply Chain Management some months before, found a private consultancy company that was fruitless from the start on, and was looking for new challenges. Johan told about this new buzzword in higher education - learning analytics. Maybe I was interested in researching its usefulness for Hogeschool Utrecht? Well over half a year later I started as a part-time researcher at HU with two objectives: provide answers to the questions "what is learning analytics?" and "what should HU do with it?" I met many people, built a strong network, did a lot of reading, and gathered many great ideas and applications. Funding, however, was always tricky. Fast forwards to September 2016: Johan called with the idea to file a proposal for a Ph.D. voucher. This would provide a better structure for my research activities and would guarantee four more years of funding. Johan volunteered to be the main supervisor. One minor detail: the proposal must be written and submitted within six days. This seemed a mission impossible - write a scientific sound research proposal, invite members for an advisory committee, have internal and external persons draft credentials et cetera. However, all the hard work in previous years paid off and within six days, the proposal was ready, submitted, and - after a minor revision - approved. After working two years in relative ease, the clock was now ticking for real, counting down to December 2020. Many thanks to Johan for 1) keeping in touch even when I wasn't a student anymore, 2) landing me a comfortable position of `junior researcher', 3) supporting me in all imaginable ways, and 4) most importantly, believing in me as a person and in the value of the work I did.

Being an `official' Ph.D. candidate now, the chain of command changed a bit as well. Most noticeably, Esther van der Stappen became my daily supervisor. We agreed on meeting every week to discuss the progress of all my work and we did so for four years. Sometimes these meetings felt painful when I did not `produce' something useful to discuss but, to be honest, this is the only way for me to get started. However, even when I had written some text or made a presentation of some sort, these meetings could feel uncomfortable. Esther is one of the sharpest and quick-thinking persons I know and can riddle any great piece of text with holes within minutes. This can be confronting at times, but it also provides great learning experiences. I owe many thanks to Esther for showing me how to conduct good research and supporting me all the way. Without her guidance and help, my Ph.D. research would probably have been crashed and burned, probably even before it was halfway.

Speaking of co-supervisors; Rogier van de Wetering became part of the team in the later stages of this research. Many thank Rogier for volunteering and making valuable contributions to this work. This is probably the shortest and – hopefully - easiest Ph.D. research you have ever been involved in.

About 25% of the time in a week should be dedicated to work, the rest is called `private time'. This, however, is on average. Working days of 12+ hours have occurred. As the balance between working life and private life is a zero-sum game, in practice this meant that my private life suffered at times. Luckily, I am blessed with a loving, understanding wife who never commented on odd working hours or me being away for conferences or other gatherings. Thank you, Mandy for being patient with me and supporting me whenever necessary. Even after our son Evelio was born, trips abroad were no problem - at least not for you. I tried to be home again as fast as possible. Fatherhood surpasses spending some extra days at nice and wonderful conference venues. I also owe many thanks to my parents: Theo and Anneke. My choices regarding education might not always have been the most optimal. Nonetheless, my parents supported me and probably knew that someday, all would be fine. Finally, that day has come.

During the years, I had the opportunity to meet a lot of people, many of them helping me in some way. My sincere thanks for that. Huub Everaert and Ilya Zitter for taking a seat on my advisory board and giving me feedback all along the way. My colleagues at Technische Bedrijfskunde, always interested in how things were going. My (former) fellow members of the Special Interest Group - Alan Berg, Maren Scheffel, Marcel Schmitz, Albert Vlaardingerbroek, Suzanne Jacobs - with

whom I grew interest in learning analytics in the Netherlands and who always lend a hand to help with some research activities when necessary. Theo Bakker, who I met in my very first week at HU and fueled my interest in analyzing student data. Later, he helped me a couple of times in organizing research activities. The people of SURF - Karianne Vermaas, Germaine Poot, Iris Huis in ‘t Veld, Marieke de Wit - van de Vecht, Jocelyn Manderveld - often provided me with a podium to share my work and make people enthusiastic for learning analytics. Another word of thanks to all my fellow doctoral students. Be able to vent the emotions that come with the job and share both positive and negative experiences is sometimes a big relief. Carry on and enjoy, one day it will all be over.

# Contents

# PART I

## Introduction

# CHAPTER

# 1

# 1. General Introduction

*It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts.*

- Sherlock Holmes (Sir Arthur Conan Doyle)

With the fourth industrial revolution in full swing, it is not surprising that the higher educational domain looks for opportunities to reap data benefits. As navigation apps can warn us of traffic jams ahead and offer alternative routes to our destination, should it not be possible to warn students of potential problems in their study path? And as text processors can give suggestions for alternative wording and sentence structure, should we not be able to build systems that can automatically provide good, automated feedback on students' work. Big data is the new oil (Hirsch, 2014), also in education, where *learning analytics* is the process of analyzing learner data to enhance education. Various definitions of learning analytics exist:

- ``The measurement, collection, analysis, and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" (LAK, 2011)
- ``Learning Analytics is about collecting traces that learners leave behind and using those traces to improve learning" (Duval, 2012)
- ``The use, assessment, elicitation and analysis of static and dynamic information about learners and learning environments, for the near real-time modeling, prediction and optimization of learning processes, and learning environments, as well as for educational decision-making" (Ifenthaler, 2015)

Although they differ in nuances, all the above definitions describe the use of learner-generated data to enhance learning and the environment in which the learning takes place. Like other forms of data analytics, the interest in learning analytics and its use became popular in the last decade. However, the concept of learning analytics itself goes back a long time, even to the 1920s (Joksimović et al., 2019). Nonetheless, the first definition of learning analytics was coined in 2011 and research shows that the numbers of publications and citations in the learning analytics field grow exponentially from that year on (Romero & Ventura, 2020; Zhang et al., 2018). Several drivers explain this relatively recent growth of the field. For example, the digitalization of education led to an increased availability of learner data (Ferguson, 2012); learning analytics provides profound insight into student behavior and performance (Gašević et al., 2015); and the increase in demand for higher educational institutions to measure, demonstrate and improve performance (Ferguson, 2012; Tsai et al., 2019).

Education cannot be optimized with the collection of data alone. It requires a multi-phase process to turn data into useful information and action. At this point, it is important to differentiate learning analytics from other forms of analytics in the educational domain, most notable, academic analytics and educational data mining. There is a certain degree of overlap between the three forms but they all have a different focus and follow different methodologies. Academic analytics is the application of business intelligence in the educational context and, in contrast to learning analytics, emphasizes institutional, regional, and even international levels rather than the learning process (Long & Siemens, 2011). Moreover, compared to learning analytics, educational data mining places more focus on automated methods for discovery within educational data, on modeling specific constructs and the relationships between them, and on automated adaptation conducted by a computer system rather than a human-being (Siemens & Baker, 2012). Campbell and Oblinger (2007) describe five steps: capture, report, predict, act, and refine. This process relates to academic analytics. Clow (2012) builds on learning theory and describes a similar model that is to better suit the learning analytics domain's needs and goals, i.e., improve learning and the environment in which learning takes place. The process of turning learner data into pedagogical interventions is now described by the Learning Analytics Cycle (Clow, 2012) - see Figure 1. The first step starts with learners who study a course, follow a MOOC, participate in a workshop et cetera. In the second step, the learners interact with learning systems, thereby leaving digital traces that are collected. Step three regards the processing of the collected data into metrics or analytics, which are communicated via dashboards, visualizations, lists et cetera. In the fourth and final step, learners, teachers, managers, or policymakers perform interventions based on the analytics. Greller and Drachsler (2012) distinguish between *data clients* and *data subjects*. The clients are the learning analytics beneficiaries and act upon the analytics' outcomes. The subjects are those who supply the data, most often the learners. Although the clients and the subjects might be the same, this is not necessarily true in all cases. For example, teachers (clients) might use data generated by learners (subjects) to improve course materials. As learners react to these interventions (or not), the cycle starts over again. According to Clow (2012), effective learning analytics should involve at least some interventions. These come in different forms, for example, offering personalized recommendations, visualizations of learning data to better understand learning behavior or support remedial action, and personalized reports on study progress or performance to allow timely action if necessary (Wong & Li, 2020). Effects of interventions on learning include, non-exhaustively, improved study performance, higher retention or registration rate, and enhanced productivity/effectiveness in learning and teaching.

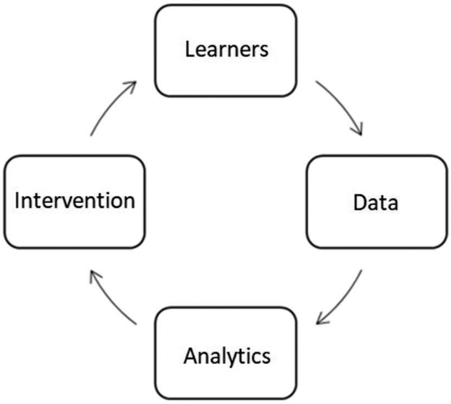


**Figure 1:** The Learning Analytics Cycle (Clow, 2012)

## 1.1. Motivation

To date, the effectiveness of learning analytics remains unclear. Although research indicates positive results (Foster & Francis, 2019), there is also research that is more cautious about the effects on learning (Francis et al., 2020). This might sound daunting but back in the days the same happened with Information Technology (IT) and, more recently, big data analytics (Gupta & George, 2016). Nonetheless, IT and big data are nowadays major pillars of modern society. In an attempt to collect evidence on the effectiveness of learning analytics, the Learning Analytics Community Exchange (LACE) project was launched (Ferguson & Clow, 2017). However, much of the learning analytics literature does not report on empirical results nor provides evidence on whether learning analytics has lived up to its promise to enhance education. As a result, there is yet no proof that the large-scale adoption of learning analytics benefits learners (Ifenthaler & Yau, 2020). In line with Knight et al. (2020), we define *large-scale adoption* as the adoption of learning analytics at multiple, smaller sites within an institution rather than focusing on learning in large-scale settings, i.e., one large group of learners. To further research learning analytics' effectiveness, wider adoption of learning analytics systems is necessary. Although learning analytics has the potential for addressing learning and teaching challenges, the systematic adoption by higher educational institutions remains low (Gašević et al., 2019). Herein lies the paradox: without evidence on its effectiveness, higher educational institutions are resistant to invest in learning analytics but without large-scale adoption, the effectiveness cannot be researched and proven or falsified. The limited number of studies that empirically validate the impact of interventions on learning is one of the challenges the higher educational institutions face that take the initiative to adopt learning analytics (Tsai & Gašević, 2017b). Even when institutions are willing to implement learning analytics, many obstacles have to be overcome as possessing the right data alone is not enough. For example, there must also be a data-informed culture and strong leadership (Macfadyen & Dawson, 2012).



To support higher educational institutions in their quest towards the successful implementation of learning analytics, several models and frameworks have been developed over the years. A distinction can be made between *input models*, *output models*, and *process models* (Broos et al., 2020; Colvin et al., 2017; Dawson et al., 2018). The first type - input models - describe dimensions that together influence the adoption of learning analytics at an institution. For example, Greller and Drachsler (2012) present a learning analytics framework with six critical dimensions, i.e., stakeholders, objective, data, instrument, external limitation, and internal limitation. Each of these dimensions requires attention and needs to be instantiated in a fully formulated learning analytics design. Bichsel (2012) developed a maturity model with five factors that require attention - culture & process, investment, expertise, governance & infrastructure, and finally data, reporting & tools. The maturity model can be used to assess the current state of analytics endeavors at higher educational institutions. A comparable input model by Norris and Baer (2013) describes five factors for organizational capacity for analytics found within leading educational institutions: technology & infrastructure, processes & practices, culture & behaviors, skills & values, and leadership. Yet another model with similar dimensions is the Learning Analytics Readiness Instrument (Arnold, Lynch, et al., 2014). This model aims at identifying deficiencies in important dimensions and providing actions to mitigate or remediate those areas. The readiness instrument comprises four dimensions - ability, data, culture & process, and governance & infrastructure. However, a major shortcoming of these input models is the lack of understanding on how the dimensions interact with each other and their connection is a complex process (Dawson et al., 2018). A second shortcoming is that it is unclear how these models should be operationalized in practice (Broos et al., 2020).

The second type of implementation model is the outcome model (Broos et al., 2020; Colvin et al., 2017; Dawson et al., 2018). This kind of model describes the dimensions important to learning analytics adoption and measures progression over time, involving different levels of readiness and maturity, and considers expected outcomes from it. A good example of such a model is the Learning Analytics Sophistication Model (Siemens et al., 2013), which comprises five stages that increasingly impact the organization as the maturity of learning analytics deployments grows. Although helping to map learning analytics outcomes, output models have downsides as well. They often do not identify all necessary dimensions to achieve these outcomes and raise the suggestion that learning analytics implementation is a linear process (Colvin et al., 2017), fail to address the complexity of learning analytics implementation (Dawson et al., 2018), and do not deliver specific guidelines for concrete projects (Broos et al., 2020).

The third type of implementation model is the process model (Broos et al., 2020; Colvin et al., 2017; Dawson et al., 2018). Compared with the previous two types of models, the process model is the one most recently emerging. It focuses on how to implement a learning analytics program by sequentially mapping steps that need to be taken. This kind of model often regards implementation as an iterative and continuous process rather than a purely linear process and can deal with the complexity of learning analytics adoption. One of the first process models is the RAPID Outcomes Mapping Approach (ROMA) framework (Ferguson et al., 2014). The framework involves seven steps, starting with defining a set of policy objectives and finishing with the development of a monitoring and learning system for evaluation. The Supporting Higher Education to Integrate Learning Analytics (SHEILA) framework is adapted from the ROMA framework and focuses on European higher educational institutions (Tsai et al., 2018). Based on the seven steps of the ROMA framework, the SHEILA framework distinguishes between six dimensions. Per dimension, it not only provides policy questions but also identifies potential challenges and strategic action points. Lately, the SHEILA framework is adapted to the needs of higher educational institutions in Latin America, resulting in the LALA framework (Broos et al., 2020). This framework considers four phases that must be passed to achieve adoption at scale: initializing, prototyping, piloting, and scaling. Although the ROMA, SHEILA, and LALA frameworks offer great support to higher educational institutions willing to implement learning analytics at scale, they also have their shortcomings. Even though the ROMA framework was used in two cases, there is still little evidence of its validity (Dawson et al., 2018). The same is true for the LALA framework, which is only validated with experts but yet lacks empirical testing (Broos et al., 2020).

On the other hand, the SHEILA framework does not always provide solutions to identified challenges (Tsai et al., 2018). Like input models, the process models lack validation and practical guidelines on how to operationalize them at scale remain scarce (Broos et al., 2020). Table 1 provides a summary of the existing models and the strengths and weaknesses of the three different model types.

| Model type | References | Strengths | Weaknesses |
|---|---|---|---|
| Input | · Greller & Drachsler, 2012 (LA Framework)<br>· Bichsel, 2012 (Maturity model)<br>· Norris & Baer, 2013 (Organizational capacity)<br>· Arnold et al., 2014 (LARI) | · A quick insight into what dimensions are important<br>· Consider a variety of dimensions, not only data/technology | · Unclear how to be operationalized<br>· Little attention for interaction between dimensions<br>· Often abstract and generic |
| Output | · Siemens et al., 2013 (LA Sophistication Model) | · Considers the desired outcomes<br>· Map development over time | · Barely addresses the complexity of implementation<br>· Do not deliver specific guidelines |
| Process | · Ferguson et al., 2014 (ROMA)<br>· Tsai et al., 2018 (SHEILA)<br>· Broos et al., 2020 (LALA) | · View implementation as an iterative and continuous process<br>· Capable of dealing with the complexity of implementation | · Empirical validation is yet low<br>· No practical guidelines on how to be operationalized |

**Table 1:** Comparison of Existing Models

From a geographical perspective, research towards learning analytics adoption mainly focuses on North America, the United Kingdom, and Australia (Ferguson & Clow, 2017; Yau & Ifenthaler, 2020). This is reflected in the research discussed before: cases can often be found in the United States (Arnold, Lynch, et al., 2014; Bichsel, 2012; Norris & Baer, 2013; Siemens et al., 2013), the United Kingdom (Ferguson et al., 2014; Siemens et al., 2013), and Australia (Colvin et al., 2015; Dawson et al., 2018; Ferguson et al., 2014; Siemens et al., 2013). Exceptions are the LALA framework (Broos et al., 2020) and the SHEILA framework (Tsai et al., 2018), which focus on Latin America and Europe. In Europe, it can be noticed that research towards learning analytics has not led to national adoption or European-level strategies (Nouri et al., 2019). Similar findings are presented by (Tsai & Gašević, 2017a): 9 out of 51 interviewed institutions had rolled out learning analytics institution-wide and only 2 out of 46 surveyed institutions did so. The same can be noticed in the Netherlands. Research findings indicate no serious threats for learning analytics implementation in Dutch higher educational institutions (Drachsler et al., 2014), and Dutch institutions indeed started to adopt learning analytics (Nouri et al., 2019). Nonetheless, Dutch institutions often cannot utilize the potential of learner data at scale yet (van der Spek, 2018; Vereniging van Universiteiten et al., 2017). As national educational systems and culture differ, insights from extant learning analytics implementation models are only useful to a certain degree to Dutch higher educational institutions willing to adopt learning analytics. As a teacher at a Dutch university of applied sciences himself, the Ph.D. researcher wants to help Dutch institutions and better understand the current barriers that prevent large-scale learning analytics adoption. This calls for a model that is specific for the Dutch context. Focusing on the Netherlands, the researcher's existing network with, among others, students, teachers, policymakers, managers, researchers, and learning analytics experts can be utilized. This is convenient, as time and resources are limited, and using exiting contacts helps to accelerate research activities.

One of the main drivers of the emergence of learning analytics is the digitalization of education (Ferguson, 2012). Information Systems (IS) play an important role in this digitization (Nguyen et al., 2020). Within IS research, the resource-based view has been used to relate IS/IT resources to organizational benefit (Cosic et al., 2015). In the resource-based view, an organization's performance is attributed to its ability to leverage resources (Barney, 1991). These resources can relate to physical capital resources (technology, equipment, raw materials, et cetera), human capital resources (training, experience, intelligence, et cetera of individual employees), and organizational capital resources (reporting structure, planning, coordinating systems, et cetera). A sustained competitive advantage can be gained when an organization possesses resources that are valuable, rare, inimitable, and non-substitutable. From a higher educational institution's point of view, the analysis of learner data can lead to improved learning, which in turn is a competitive advantage. Although higher educational institutions in the Netherlands nowadays collaborate to a certain degree, the rivalry between institutions will likely not disappear entirely (HOP et al., 2020). Therefore, being able to provide better education to learners is an important strategic point. Also, educational institutions have the lawful and moral obligation to provide the best education they can offer. However, only possessing the right resources alone is not enough to gain sustainable competitive advantages, as a productive activity requires cooperation and coordination between resources (Grant, 1991). Capabilities play a paramount role here. Capabilities can be defined as:

- ``[T]he ability for a team of resources to perform some task or activity." (Grant, 1991, p. 119)
- ``[A] firm's capacity to deploy *resources*, usually in combination, using organizational processes, to effect a desired end." (Amit & Schoemaker, 1993, p. 35)
- ``[A]n organizationally embedded nontransferable firm-specific resource whose purpose is to improve the productivity of the *other* resources possessed by the firm." (Makadok, 2001, p. 389)

The definitions show that capabilities are a special kind of resource and describe how resources should interact to benefit an organization. In contrast to normal resources, capabilities are non-transferable and must be developed by an organization itself (Makadok, 2001). They transform inputs into outputs of greater worth and include skills and processes (Wade & Hulland, 2004). However, capabilities can only turn inputs into outputs after the other resources have been acquired (Makadok, 2001). This aligns with the observation that learning analytics can only improve learning after it has been adopted at a large scale within a higher educational institution - without learner data, skillful personnel, and analytical software, there is little use for effective data-informed processes. By reviewing adjacent research fields, we can see that several capability models for analytics exist. Cosic et al. (2015) developed a capability model for business analytics. It comprises 16 capabilities distributed over four categories: governance, culture, technology, and people. Moreover, Gupta & George (2016) describe three groups of resources (tangible, human, and intangible) that are needed to create big data analytics capabilities. Looking at the learning analytics domain, we can conclude that capabilities prove important to turn learner data into effective pedagogical interventions but to date, no capability model for learning analytics exist. As described before, extant models support the adoption of learning analytics but none of them is grounded in the resource-based view nor consider the importance of capability development.

Based on the above analysis of existing literature, we can identify several knowledge gaps in the current learning analytics domain:

- The large-scale adoption of learning analytics by higher educational institutes is yet low (Gašević et al., 2019). As a result, there is little evidence of learning analytics' effectiveness (Ifenthaler & Yau, 2020).
- The adoption of learning analytics is a complex undertaking. To support higher educational institutes in this challenge, various models have been developed (Broos et al., 2020; Colvin et al., 2017; Dawson et al., 2018). However, these models have their shortcomings. Most notable, the way how to operationalize them at scale remains unclear.
- Process models are regarded as the most sophisticated model type and best suit the complexity of learning analytics adoption (Broos et al., 2020; Colvin et al., 2017; Dawson et al., 2018). However, empirical validation of these models at a wide scale and practical guidance is still missing.
- Initially, research on the adoption of learning analytics focuses on North America, the United Kingdom, and Australia (Ferguson & Clow, 2017; Yau & Ifenthaler, 2020). Models for the European mainland and Latin America were developed only recently. To make models work in a context other than the one it was designed for, adaption is necessary.
- Capabilities are an important kind of resource for higher educational institutions to turn learner data into effective pedagogical interventions. Nonetheless, there is no capability model for the learning analytics domain yet.

## 1.2. Research questions

In the previous section, we summarized several gaps in the current learning analytics domain. This Ph.D. research aims to design a capability model to support the adoption of learning analytics by higher educational institutions in the Netherlands. The main research question is as follows:

"*Which Information System-related capabilities for learning analytics benefit teachers and learners in Dutch higher educational institutions?*"

As described in paragraph 1.3, this research follows the Information Systems Design Science Research approach presented by Hevner et al. (2004). This thesis is structured accordingly and comprises four parts. Each part answers one or more research questions that are derived from the main research question. The structure is as follows:

- Part I: Introduction
- Part II: Rationale

- Part III: Design
- Part IV: Evaluation

We will now elaborate on each part and the research questions they cover.

### 1.2.1 Part I: Introduction

The goal of learning analytics is to improve learning and learning environment (LAK, 2011). Therefore, the implementation of learning analytics at higher educational institutions is successful when it achieves this goal. The question is how to measure the beneficial effect of learning analytics on learning. Often, visible and measurable targets such as grades and retention are used (Prakash et al., 2014). However, these measures only capture learning outcomes, not the learning process itself or the learning environment. To be able to measure all outcomes of learning analytics, the following research question is answered in chapter 2:

*"In what way does existing literature on learning analytics interventions operationalize affected learning?"*

### 1.2.1. Part II: Rationale

The second part of this Ph.D. thesis shows the rationale of the research. Higher educational institutes face many challenges when implementing learning analytics (Tsai & Gašević, 2017b). To better understand these challenges and experience them in a real-life context, chapter 3 provides an answer to the following research question:

*"What issues are encountered when implementing an experimental learning analytics tool in the case organization's virtual learning environment?"*

Data are at the learning analytics' core (LAK, 2011) and considered an important, tangible resource for organizations across all industries (Gupta & George, 2016). However, just possessing data does not lead to any improvement. Data must be collected, analyzed, visualized, and used for intervention to enhance learning (Clow, 2012). Although this sounds simple, it is a tedious process. That is, raw data exported from learning systems need to be cleaned and transformed before it is of any use to educators and students. It is estimated that this task takes up to 80% of the analytical time (Brink et al., 2016). To research what capabilities are important to this part of the learning analytics process as show the effects on its outcomes, chapter 4 answers the research question:

*"What are the effects of (unspoken) choices made during the cleaning process of student data on the outcomes when these data are in turn used for learning analytics?"*

### 1.2.2. Part III: Design

To the best of our knowledge, the resource-based view is not used to support the adoption of learning analytics yet. In adjacent domains like business analytics (Cosic et al., 2015) and big data analytics (Gupta & George, 2016), resource-based capability models help organizations to apply successful data analytics. The aim of chapter 5 is to design a capability model for the learning analytics domain based on theory from the extant knowledge base:

*"What capabilities for the successful adoption of learning analytics can be identified in existing literature on big data analytics, business analytics, and learning analytics?"*

Chapter 5 leads to the first, theoretical version of the Learning Analytics Capability Model. After its construction, a model needs to be evaluated and the feedback from this evaluation leads to the model's refinement (Hevner, 2007). Chapter 6 elaborates on these steps and provides an answer to the following question:

*"How can the Learning Analytics Capability Model be evaluated and refined based on empirical data from a single Dutch higher educational institution that is mature in the use of learner data to improve learning?"*

### 1.2.3. Part VI: Evaluation

Any design science research project should end with an evaluation of the designed artifact to assures its rigor (Venable et al., 2016). Chapter 7 of this Ph.D. thesis describes the evaluation process and provides an answer to the following question:

*"How to ex-post evaluate the Learning Analytics Capability Model in the context of Dutch-speaking educational institutions?"*

## 1.3. Research design

To answer the main research question of this Ph.D. research, we will design a capability model: the Learning Analytics Capability Model. As Information Systems are important to learning analytics (Ferguson, 2012; Nguyen et al., 2020), we opt to use the Information Systems Design Science Research approach presented by Hevner et al. (2004). Information Systems Design Science principles are applicable when solutions to so-called ‘wicked problems’ are required. We argue that implementing learning analytics is indeed a wicked problem since the interaction between subcomponents of the problem and its solution are complex and the multidisciplinary nature of learning analytics calls for well-developed human social abilities. A second reason to use Hevner et al.’s research framework is that during this Ph.D. research, an IS/IT artifact is developed. Among others, IS/IT artifacts relate to models, i.e., abstractions and representations (Hevner et al., 2004). Hevner et al. (2004) provide seven guidelines for design science research. Although one should be careful with mechanically applying all these guidelines (Venable, 2010), they provide useful suggestions on how to structure a design science project. Therefore, we apply these guidelines to structure the Ph.D. research at hand. Hevner (2007) distinguishes three research cycles that must be present in a design science research project: the relevance cycle, the rigor cycle, and the design cycle. The relevance cycle initiates the research by providing the requirements as well as the criteria for the evaluation of the research outcomes. To assure the research’s relevance, research activities must address practice needs. These needs come from an environment that is composed of people, organizations, and technology. We conducted a single-case study to analyze the issues faced by a higher educational institution when implementing an experimental learning analytics environment (chapter 3). Moreover, to highlight the importance to have well-developed capabilities, we conducted a detailed study on the effects of data quality on learning analytics outcomes (chapter 4). The second cycle described by Hevner (2007) - the rigor cycle - ensures the research’s innovativeness. To assure the research’s rigor, the existing knowledge base must be consulted. The knowledge base comprises foundations (existing theories, frameworks et cetera) and methodologies. During a systematic literature review, we reviewed what capabilities for data analytics can be identified in extant literature (chapter 5). Since the learning analytics research domain is relatively young, we applied the exaptation process (Gregor & Hevner, 2013), i.e., review literature from the business analytics and big data analytics domains for known capabilities important to analytics. The results of the systematic review literature served as input for the third and last cycle described by Hevner (2007) – the design cycle. The design cycle lies at the heart of the research project and involves the construction and refinement of the artifact. Based on the systematic review literature, the first, theoretical Learning Analytics Capability Model was designed. After initial development, the research outcomes must be justified and evaluated. This in general leads to refinement and reassessment. The theoretical model is refined using empirical data collected during a single-case study at a higher educational institution that is mature in the use of learner data to enhance education (chapter 6). This study led to the refined Learning Analytics Capability Model. Finally, to study and evaluate the artifact's use in the application domain, it must be returned into the environment (Hevner et al., 2004). In a mixed-method study that comprised pluralistic walkthroughs, expert evaluation, and a survey, the model is empirically evaluated and validated by practitioners from the application domain (chapter 7).

## 1.4. Research method

The studies conducted in this Ph.D. research use a mix of research methods. As all but one method are qualitative, we argue that our research follows a multimethod research approach (Venkatesh et al., 2013), which leads to rich and reliable research results (Mingers, 2001). The different methods that are applied during the various studies are shown in Table 2. Next, we briefly describe each method.

| Research method | Chapter 2 | Chapter 3 | Chapter 4 | Chapter 5 | Chapter 6 | Chapter 7 |
|---|---|---|---|---|---|---|
| Narrative review | | X | X | | X | X |
| Systematic literature review | X | | | X | | |
| Case study | | X | X | | X | |
| Pluralistic walkthrough | | | | | | X |
| Group discussion | | | | | | X |
| Survey | | | | | | X |

**Table 2:** Research Methods per Study

### 1.4.1. Literature review

To explore what is already known about the topic at hand, each study started with a literature review. This can be classified as a *narrative review*, as the goal is to identify what has been written on a subject rather than generalizing or accumulating knowledge (Paré et al., 2015). As often the case with narrative reviews, no explanations on how the review process is conducted are provided in the studies' resultant papers. In general, the reviews started opportunistic, i.e., with literature that the researchers already knew. Next, key search terms were used to find potentiality interesting literature via Google Scholar, ResearchGate, and other scientific databases. Once a set with relevant research papers was established, a snowballing technique was applied to find additional information via backward and forward searches.



The studies described in chapters 2 and 5 went a step further. *Systematic literature reviews* were conducted in these two studies. Extant empirical studies were used to describe higher orders of theoretical structures. Therefore, the reviews conducted in chapters 2 and 5 can be regarded as theoretical reviews (Paré et al., 2015). The goal of the systematic literature reviews was to explorer the existing knowledge base, ensuring the Ph.D. research's rigor (Hevner et al., 2004).

### 1.4.2. Case study

Case studies allow researching contemporary phenomena that are hard to study in isolation (Runeson & Höst, 2009) and is a suitable method when no control over behavioral events is required (Yin, 2013). Case studies are common in the Information Systems research domain (Runeson & Höst, 2009) and we applied them in three cases. In chapter 3, a case study was used to research what issues are encountered when a learning analytics system is implemented. In chapter 4, a case study was conducted to assess the effects of choices made during the data cleaning process are on the analytical outcomes. Finally, in chapter 6 a case study is used to research what resource-based capabilities are present at a Dutch higher educational institution that is mature in the use of learner data to improve education.

### 1.4.3. Pluralistic walkthrough

In the study described in chapter 7, pluralistic walkthroughs are utilized to ex-post evaluate the Learning Analytics Capability Model with real users in a real context. In Information Systems research, pluralistic walkthroughs are often conducted to evaluate user interfaces but can also be used to validate other sorts of evaluands (Dahlberg, 2003; Emaus et al., 2010; Kusters & Versendaal, 2013). Pluralistic walkthroughs are defined by five characteristics (Bias, 1994; Riihiaho, 2002):

- Three types of participants are present: the designers of the system, usability experts, and users.

- (Hard-copy) panels present the system.
- All participants act like users.
- Participants write down actions performed to complete given tasks.
- After the walkthrough, there is a group discussion.

In line with Prat et al.'s (2015) suggestion to creatively and pragmatically generate new evaluation methods, we developed an evaluation method based on the above characteristics. One benefit of pluralistic walkthroughs is their speed, as generates immediate feedback from users (Thorvald et al., 2015). Another benefit is the involvement of actual users. This aligns with the need to have important stakeholders involved in the learning analytics implementation process (Hilliger et al., 2020).

### 1.4.4. Group discussion

A group discussion is conducted during the study described in chapter 7. Although the name might imply otherwise, our group discussion bears more similarity with a group interview than with a focus group discussion. That is, the researcher asked questions, controlled the discussion's dynamics of the discussion, and at times engaged in dialogues with a specific individual. This is characteristic for group interviews and less for focus group discussions, where the researcher adopts a role as facilitator (Nyumba et al., 2018).

### 1.4.5. Survey

In the evaluation study described in chapter 7, a survey in the form of a self-completion questionnaire is conducted. While the researcher himself was present during the other two methods used in this study, this was not the case during the survey. The absence of interviewer effects is therefore one of the benefits of a self-completion questionnaire (Bryman & Bell, 2015). The questionnaire is used to measure the participants' overall opinion of the Learning Analytics Capability Model, its perceived usefulness, and its ease-of-use. Within the learning analytics domain, a similar approach is used by other scholars (Ali et al., 2013; Rienties et al., 2018).

## 1.5. Thesis outline

The Ph.D. thesis comprises four parts, each featuring one or more chapters. We now briefly present each chapter.

### 1.5.1. Part I: Introduction

Chapter 1 presents the overall introduction of the Ph.D. thesis. It describes the research's motivation, the research questions that are answered in the various studies conducted during the research, the used research design, and elaborated on the applied research methods.

Chapter 2 provides operational definitions to evaluate the beneficial effects of learning analytics interventions on learning and learning environments. During a systematic literature review, 62 key studies were analyzed to identify measures of affected learning. The found operational definitions were classified in one of three distinct categories: learning environment, learning process, and learning outcome. By synthesizing the results, 11 subcategories could be distinguished. As described in later chapters, mechanisms to measure the effects of learning analytics interventions are one of the capabilities that must be developed by higher educational institutions that implement learning analytics at scale. *Chapter 2 has been published in IEEE Transactions on Learning Technologies* (Knobbout & van der Stappen, 2020b). *An earlier version has been published in the proceedings of the 13th European Conference for Technology Enhanced Learning (EC-TEL)* (Knobbout & van der Stappen, 2018).

1

### 1.5.2. Part II: Rationale

In Chapter 3, we conducted a single-case study and researched what difficulties were faced by a Dutch higher educational institution that implements a learning analytics system. Problems related to five dimensions: stakeholders, data, instruments, external limitations, and internal limitations. The study showed that implementing learning analytics is not a trivial task and many dimensions need careful consideration, providing a practical trigger and relevance for the research according to the Design Science Research approach. *Chapter 3 has been published in the proceedings of the 30th Bled eConference* (Knobbout & van der Stappen, 2017).

Chapter 4 highlights the importance of possessing the right organizational capabilities regarding learner data. During a case study, pre-formulated pedagogical questions could not be answered due to a large number of missing data. Choices had to be made to solve this issue. However, the choices that are made during the data cleaning process have consequences on the outcomes. We showed this with a case where data was gathered during six courses taught via Moodle. We illustrate possible choices in dealing with missing data by applying the cleaning process twelve times with different choices on copies of the raw data. Consequently, the analyses' outcomes show significant differences. *Chapter 4 has been published in the proceedings of the 32$^{th}$ Bled eConference* (Knobbout et al., 2019).

### 1.5.3. Part III: Design

Chapter 5 describes the Learning Analytics Capability Model's design process. 15 key studies from research domains such as business analytics, big data analytics, and learning analytics were identified during a systematic literature review. By coding the 461 operational definitions provided by the key studies, a capability model with 34 capabilities was designed. Capabilities were classified into one of five categories: *Data, Management, People, Technology*, and *Privacy & Ethics. Chapter 5 has been published in the International Journal of Learning Analytics and Artificial Intelligence for Education (iJAI)* (Knobbout & van der Stappen, 2020a).

Chapter 6 follows up on the results of the previous chapter. By conducting a single-case study, the theoretical model is evaluated and refined with empirical data. Data was collected by interviewing four stakeholders at a Dutch higher educational institution that is mature in learner data usage to improve education. As a result, the Learning Analytics Capability Model was refined: we merged seven capabilities, renamed three others, and improved all capabilities' definitions. *Chapter 6 has been published in the proceedings of the 26th Americas Conference on Information Systems (AMCIS)* (Knobbout et al., 2020).

### 1.5.4. Part IV: Evaluation

Chapter 7 describes the ex-post evaluation of the Learning Analytics Capability Model. During five pluralistic walkthroughs at five different educational institutions in the Netherlands and Belgium, 26 practitioners used the model to plan the (further) implementation of learning analytics at their institution. Moreover, a group discussion with seven learning analytics experts was held to evaluate the model. The outcomes of the pluralistic walkthroughs, the group discussion, and an individual survey showed the model's usefulness and completeness. *Chapter 7 has been submitted for review* (Knobbout et al., 2021).

Chapter 8 provides the general discussion of this Ph.D. thesis. It elaborates on the results and outcomes of the various studies conducted during the research, the contributions and implications to both academics and practitioners, the limitations, and presents directions for future research.

The overall cohesion between the chapters according to the Design Science Research framework of Hevner et al. (2004) is shown in Figure 2, where each number represents a chapter in this thesis. The general introduction and general discussion are added to provide an inclusive overview of the entire thesis.



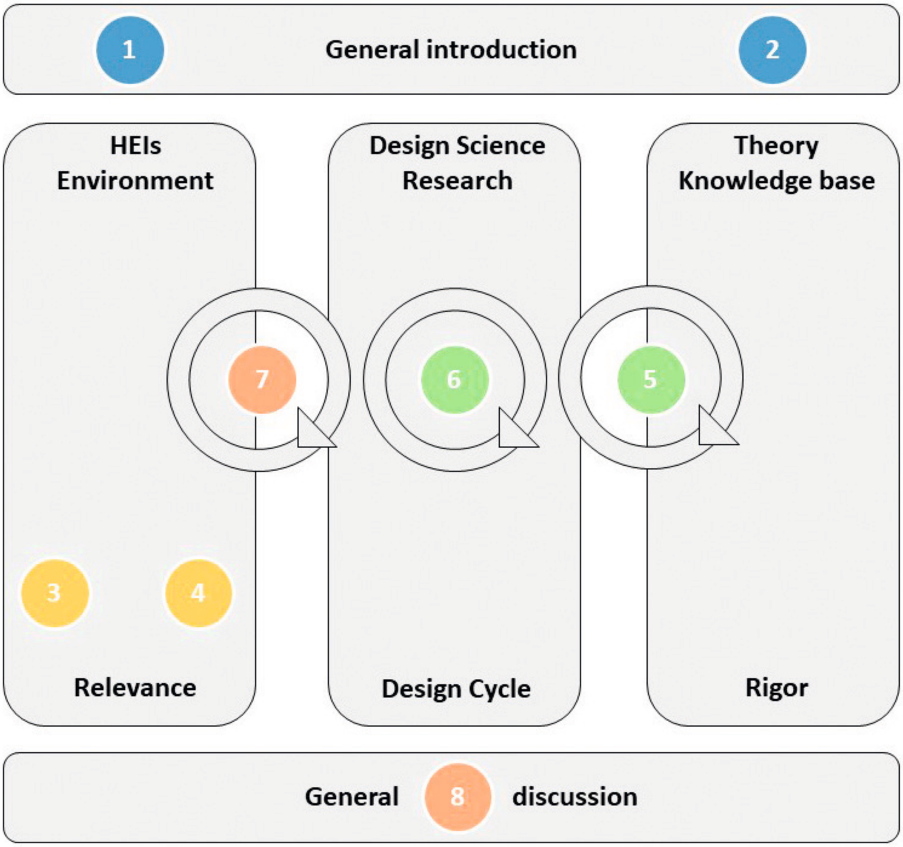


**Figure 2:** Ph.D. Thesis Overview

# CHAPTER

# 2

# 2. Where is the Learning in Learning Analytics? A Systematic Literature Review on the Operationalization of Learning-Related Constructs in the Evaluation of Learning Analytics Interventions[1]

*Learning technologies enable interventions in the learning process aiming to improve learning. Learning analytics provides such interventions based on analysis of learner data, which are believed to have beneficial effects on both learning and the learning environment. Literature reporting on the effects of learning analytics interventions on learning allows us to assess in what way learning analytics improves learning. No standard set of operational definitions for learning affected by learning analytics interventions is available. We performed a systematic literature review of 1932 search hits, which yielded 62 key studies. We analyzed how affected learning was operationalized in these key studies and classified operational definitions into three categories: learning environment, learning process and learning outcome. A deepening analysis yielded a refined classification scheme with eleven subcategories. Most of the analyzed studies relate to either learning outcome or learning process. Only nine of the key studies relate to more than one category. Given the complex nature of applying learning analytics interventions in practice, measuring the effects on a wider spectrum of aspects can give more insight into the workings of learning analytics interventions on the different actors, processes and outcomes involved. Based on the results of our review, we recommend making deliberate decisions on the (multiple) aspects of learning one tries to improve by applying learning analytics. Our refined classification with examples of operational definitions may help both academics and practitioners doing so, as it allows for a more structured, grounded and comparable positioning of learning analytics benefits.*

---

1 This work was originally published as: Knobbout, J.H., and van der Stappen, E.J. (2020). Where is the Learning in Learning Analytics? A Systematic Literature Review on the Operationalization of Learning-Related Constructs in the Evaluation of Learning Analytics Interventions, in *IEEE Transactions on Learning Technologies*, doi: 10.1109/TLT.2020.2999970.
Prior work was originally published as: Knobbout, J.H., & van der Stappen, E.J. (2018). Where Is the Learning in Learning Analytics? A Systematic Literature Review to Identify Measures of Affected Learning. In *ECTEL 2018 Proceedings*. doi: 10.1007/978-3-319-98572-5_7

## 2.1. Introduction

Learning technologies enable interventions in the learning process aiming to improve learning. Whenever such technologies are based on analytics of data on learners, the learning process and/or the learning environment, we speak of learning analytics.

Learning analytics is commonly defined as ‘the measurement, collection, analysis, and reporting of data about learners and their contexts, for the purposes of understanding and optimizing learning and the environments in which it occurs’ (LAK, 2011). Interventions in the learning process based on data from that process are believed to have beneficial effects on learning and the learning environment. These interventions are an important step that ‘closes the loop’ in the cyclic learning analytics process (Clow, 2012).



The largest challenge for learning analytics research and practice is to find out which types of interventions have a positive impact on learning (Rienties et al., 2017). In his comprehensive book on the field of learning analytics, Sclater (2017) dedicates a chapter to interventions, with a focus on human-mediated interventions taken directly with learners while learning is taking place. He concludes that there is relatively little knowledge on how these interventions can be performed effectively, even though it is a vital part of the process to provide analytics that enable actions with a beneficial impact on learners (Sclater, 2017).

These beneficial effects are increasingly subject of study (Ferguson & Clow, 2017; Gašević et al., 2015; Larrabee Sønderlund et al., 2018; Mangaroska & Giannakos, 2019; Viberg et al., 2018). These studies all support that 1) there are relatively few studies that report on human-mediated interventions taking place directly in the learning process (at the micro-level), and 2) there is little evidence available on the desired improvement of learning. Several recent studies call for more (longitudinal) empirical research in authentic settings as well as for a more systematic comparison of learning analytics interventions (Ferguson & Clow, 2017; Larrabee Sønderlund et al., 2018; Schwendimann et al., 2017; Viberg et al., 2018). To systematically design, implement and evaluate learning analytics interventions, it is important to know how to measure the intended improvement of learning. Central to our study is the concept of ‘affected learning’, which denotes the observable change in learning caused by learning analytics interventions. A shared, transparent and tested set of operational definitions for learning affected by learning analytics interventions is also crucial to enable comparison and generalization of studies on learning analytics interventions - and learning technologies in general - and eventually meta-studies on effect sizes.

In this paper, we study what operational definitions for affected learning can be identified in existing literature on learning analytics interventions. We conduct a systematic literature review in order to provide an answer to the research question: In what way does existing literature on learning analytics interventions operationalize affected learning?

We structure the results of our study based on a high-level, preliminary classification scheme derived from learning theory. This classification scheme is subsequently refined in the analysis phase of the review. Our research supports both academics and practitioners in their work as it provides (1) a refined classification scheme for operationalizing affected learning and (2) actual operational definitions of affected learning which can be used to measure and compare the intended benefits of learning analytics interventions on learning. We structure the remainder of this paper as follows. First, we provide an overview of the background of the study and related reviews from the field. We then describe the methodology, followed by an elaboration on the analysis and results. Finally, we provide recommendations for future research and discuss the limitations of our study.

## 2.2. Background

In this section, we give an overview of learning analytics interventions and recent reviews of the learning analytics field. Furthermore, we introduce a preliminary classification scheme which we use for the analysis in this review.

### 2.2.1. Learning analytics and learning analytics interventions

The Learning Analytics Cycle (Clow, 2012) describes the process of turning data into action and involves four steps: 1) learners generate data, 2) the infrastructure captures, collects and stores this data, 3) the collected data are analyzed and visualized, and 4) feeding back this analytics and/or visualizations to stakeholders such as learners and teachers. Such a learning analytics intervention is needed in order for learning analytics to have effect on learners. Learning analytics interventions can be defined as “the surrounding frame of activity through which analytic tools, data, and reports are taken up and used” (Wise, 2014).

These interventions can have a wide variety of appearances, e.g., automated visualizations of students' progress in the form of learning analytics dashboards (Schwendimann et al., 2017), early warning systems for educators to identify students at risk (Pistilli & Arnold, 2012), supporting adaptive learning pathways (Mavroudi et al., 2018), and goal-setting recommendations based on labor market data analytics (Kobayashi et al., 2014). Learning analytics interventions are not restricted to formal educational settings, however, the number of studies in the context of non-formal or workplace learning is very limited (Ruiz-Calleja et al., 2017; Schwendimann et al., 2017). Examples of changes in the learning process achieved by these interventions are personalization of learning, enhanced instructor support of learners or improvement of curricula (Nunn et al., 2016; Sclater, 2017). The effectiveness of learning analytics can be enhanced by increasing the speed of delivery of learning analytics interventions (e.g. real-time feedback to learners and teachers) (Clow, 2012).



In the definition of learning analytics, the goal described is twofold: we aim to understand and optimize learning (LAK, 2011). Learning analytics takes place at the micro-level within educational institutes, so the focus is on the learner and its surroundings (Barneveld et al., 2012). In this study, we focus on studies in which 1) learning analytics interventions have been performed 2) in authentic settings and 3) empirically evaluated with respect to learning-related constructs. Studies in which, for instance, (advanced) analytics have been performed on learning data to measure the effects of a new instructional strategy or course design do not fall into our scope, since no 'data-driven intervention' is performed.

### 2.2.2. Recent reviews

In recent years, the number of (systematic) reviews of the learning analytics field is increasing rapidly. One of the first systematic reviews of learning analytics literature classifies studies by learning setting, analysis method and research objectives (Papamitsiou & Economides, 2014). That study shows that learning analytics uses a wide variety of techniques and is not limited to only Virtual Learning Environments (VLEs), but can also be applied on, among others, web-based education, social learning, and cognitive tutors. The objectives of the studies are diverse and include e.g., student behavior modelling, prediction of performance, prediction of dropout and retention, recommendation of resources, and increased (self-) reflection and (self-) awareness.

The learning analytics field is relatively young but steadily maturing, which is also noticeable in the increasing attention that is given to the evidence-based character of the field. Recently, (Ferguson & Clow, 2017) analyzed the evidence in the LACE Evidence Hub and they conclude that there is considerable scope for improving the evidence base for learning analytics. Among other aspects, they suggest paying more attention to the cyclic nature of learning analytics (closing the cycle) and to the validity, reliability and generalizability of learning analytics research.

In 2017, Schwendimann et al. presented a systematic literature review of research on learning dashboards. Based on their review, they define a learning dashboard as ‘a single display that aggregates different indicators about learner(s), learning process(es) and/or learning context(s) into one or multiple visualizations’ (Schwendimann et al., 2017), clearly distinguishing dashboards from visualizations based on a single indicator. Interestingly, the indicators used in the dashboards in 60% of the papers included in the review were gathered from authentic educational situations, whereas merely 29% of the included studies actually evaluated the dashboard in such situations. Of all 55 analyzed papers, only four evaluated the impact of the dashboard on learning, whereas most others evaluated other aspects such as usability and user satisfaction. Based on the results from their review, the authors conclude that large-scale studies on adoption and learning impact of dashboards are important yet under-explored. Schwendimann et al. also observe a lack of comparative studies in the field, partly due to ‘a lack of widely-accepted, specific evaluation constructs, beyond general ones like usability and usefulness’ (Schwendimann et al., 2017). In our study, we aim to support the development of a set of operational definitions for the construct of affected learning.

Mangaroska and Giannakos (2019) performed a systematic literature review on how learning analytics have been used to inform learning design. They aimed to gain insights on the intersection of these two research fields rather than the individual disciplines. The authors emphasize the need for actionable insights from learning analytics, i.e., data-driven interventions fed back to stakeholders in the learning process, thereby closing the learning analytics loop effectively. Out of the 43 analyzed papers, just four reported learning analytics integrated into a learning environment providing real-time feedback. In their discussion, Mangaroska and Giannakos state that researchers should ‘know what data to collect in order to understand whether certain learning processes are activated, and what learning outcomes are associated with what design decisions’ (Mangaroska & Giannakos, 2019) and they urge learning analytics researchers to 'evaluate and denote student learning outcomes, or any other learning-related constructs' (Mangaroska & Giannakos, 2019).

Several recent systematic literature reviews focus on higher education as a specific educational context for learning analytics (Larrabee Sønderlund et al., 2018; Viberg et al., 2018). Viberg et al. (2018) performed a comprehensive review of 252 papers on learning analytics in higher education, analyzing the research approaches, methods and evidence for learning analytics. With respect to the latter, they examined evidence for propositions 1 and 2 from Ferguson and Clow (2017): 1) learning analytics improve learning outcomes and 2) learning analytics improve learning support and teaching. They found that only 9% of the studies reported on evidence for proposition 1 and 35% found evidence for proposition 2. Interestingly, they include studies in their review such as (Guarcello et al., 2017) and (Gašević et al., 2014), which both study an instructional approach (supplemental instruction and grading self-reflection video annotations, respectively) and apply advanced analytics to assess the effects of such an instructional approach rigorously. Viberg et al. (2018) include these studies as evidence for improvement of student outcomes by learning analytics, whereas they both fall outside our definition of learning analytics interventions, since the instructional interventions themselves were not based on data analytics.



Larrabee Sønderlund et al. (2018) performed a systematic literature review specifically aimed at studying the effectiveness of learning analytics interventions based on predictive models. From 689 papers, merely 11 studies reported on an evaluation of the effectiveness of such interventions. They conclude their review emphasizing the need for a solid knowledge base on the feasibility, effectiveness and generalizability of the implementation and evaluation of learning analytics interventions. In order to replicate experiments, and to compare and generalize obtained results, we need to be transparent in the (operational) variables we use to measure the impact of learning analytics interventions on learning.

Most recently, Wong and Li (2020) presented a review of 24 case studies of learning analytics interventions in higher education, analyzing objectives, data sources, intervention methods, obtained outcomes and observed challenges. The review of these case study suggests that learning analytics interventions have the potential for a broad application in terms of various purposes as well as different learning contexts within higher education. Wong and Li conclude that to fulfill the recognized potential of learning analytics interventions, 'more studies on empirical evidence, even with null or negative results, are needed to support its long-term effectiveness and sustainability' (Wong & Li, 2020).

The previously mentioned studies note the importance to evaluate the effects of learning analytics based on learning constructs. In the present study, we analyze in what way the effects on learning of learning analytics interventions (in all learning contexts) are measured by selecting key studies which report on empirical, quantitative results of the application of learning analytics interventions at the micro-level in the learning process in an authentic context. The outcome of our study is a classification scheme for constructs related to affected learning - with their operationalization - so in future research the effects of learning analytics on learning can be described comparably.

### 2.2.3. Preliminary classification of affected learning

To evaluate the effects of learning analytics interventions on learning, the difference in learning caused by the provided intervention should be measured. This raises the fundamental question in what way(s) learning can be measured. Joksimović et al. (2018) recently explored how learning is modeled in MOOC research. They present a framework specifically suitable for open online contexts with a focus on student engagement. Along similar lines, we aim to analyze how learning-related constructs are operationalized in research on learning analytics interventions in all learning contexts (K-12, higher education, MOOCs and the workplace).

Learning can either be described as a process or as the outcome of this process: a (relatively permanent) change in a person's behavior, knowledge and/or skills (Braungart et al., 2007). Not all learning theories award the same weight to both process and result. For example, the experiential learning theory by Kolb has a preference for a process-focused view: "learning is best conceived as a process, not in terms of outcomes" (Kolb, 1984). Behaviorism focuses mainly on learning outcomes, cognitivism made a shift towards taking the (cognitive) process more into account, whereas constructivism focuses mainly on the learning process (Cooper, 1993).

This process-product duality is also present in the well-established 3P model of teaching and learning (Biggs & Telfer, 1987). The framework on MOOC learning by Joksimović et al. (2018) also distinguishes process and product, while adding a third category of learning-related constructs: learning contexts. The context in which learning takes place is also present in the 3P model in the factor Presage

- Teaching context (Biggs & Telfer, 1987). We argue that learning context should be an important aspect in research on learning analytics interventions, since the most commonly used learning analytics definition states we aim to optimize not only learning itself but also 'the environments in which it occurs' (LAK, 2011).

Based on the discussion above, and in line with Joksimović et al. (2018), we now discern three categories that we will use to classify operational definitions of learning affected by learning analytics interventions: (i) Learning environment, (ii) Learning process and (iii) Learning outcome. The 3P model not only describes the factors of learning (Presage-Process-Product) but also the relations between the factors and reciprocal influences. Our high-level, preliminary classification scheme with categories, as well as their relations according to the 3P model, is shown in Figure 3. During the analysis of our review, we will refine this scheme using the identified operational definitions.



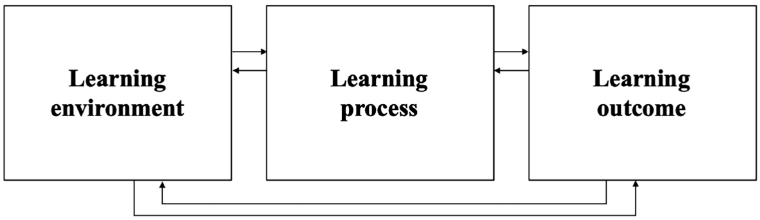


**Figure 3:** High-level, preliminary classification scheme for operational definitions of learning affected by learning analytics interventions.

## 2.3. Methodology

In this section, we provide a detailed description of the method used for our systematic literature review. The applied method in this literature review builds on other systematic literature reviews in the learning analytics domain (Bodily & Verbert, 2017; Nunn et al., 2016; Papamitsiou & Economides, 2014; Ruiz-Calleja et al., 2017). In our study, we aim to provide an answer to the following research question: *In what way does existing literature on learning analytics interventions operationalize affected learning?*

### 2.3.1. Literature sources

During the literature review, papers from seven different databases are sourced:

1. Learning Analytics and Knowledge (LAK) is the main conference in the learning analytics field. Organized for the first time in 2011, it produced an extensive

amount of proceeding papers ever since. In this study, we include the LAK conference proceeding papers;
2. IEEE Xplore is a technical-oriented database and contains papers related to, among others, computer science;
3. SpringerLink is the Springer's online collection of scientific, technological and medical journals, books and reference works;
4. The Association for Computing Machinery (ACM) database is a large, comprehensive database focused on computing and information technology;
5. ScienceDirect is Elsevier's information solution for researchers and includes over 3,800 journals;
6. The Education Resources Information Center (ERIC) database is focused on educational literature and resources;
7. Learning Analytics Community Exchange (LACE) was a European Union funded project and one of the project aims was to collect evidence of the effects learning analytics have on education. In the study at hand, we include papers which relate to the proposition "Learning analytics improve learning outcomes" (Learning Analytics Community Exchange, 9999).

### 2.3.2. Search terms

To search the aforementioned databases for literature related to operational definitions of affected learning, different search terms are used. The search terms are formulated based on a priori analysis of relevant papers. Generally, the search includes the terms "learning analytics" AND student* AND (achievement OR "student learning" OR "learning goal" OR "learning outcome" OR performance OR "student success"). When allowed for by the search engine, we specifically search the abstracts for student* and ("learning analytics") to ensure we get learning analytics-related articles.

### 2.3.3. Selection of papers and inclusion criteria

The aim of this study is to identify operational definitions of learning affected by learning analytics interventions in an authentic context. We therefore focus on quantitative studies, as they provide us with actual metrics of learning which can be calculated and compared in a standardized way. With this approach, we follow Joksimović et al. (2018). We concur with them that qualitative studies are fundamentally different in the way evaluation results are presented, which are worthy of a separate literature review, potentially yield complementary insights, e.g. thick descriptions of the constructs, variables and operational definitions we find in our review. The inclusion and exclusion criteria we used in our study are listed in Table 3.

| Inclusion criteria | Exclusion criteria |
| --- | --- |
| · Written in English<br>· Published between January 2011 and April 2019<br>· Empirical study in an authentic educational context<br>· Performs an intervention in the learning process based on learning data analytics | · Full-text not available to researchers<br>· Manuscript not peer-reviewed<br>· Theoretical study<br>· Lacks reporting of quantitative results<br>· Simulation study |

**Table 3:** Inclusion and exclusion criteria



From the papers found in the previous step, the title and abstract are read to determine whether it meets the inclusion criteria, or it should be excluded based on the exclusion criteria. Papers clearly not meeting the criteria are dismissed. If the abstract and title do not provide enough information to make the selection, the paper is scanned - especially the method and result section - to make a better-informed decision. In a second round of selection, the qualifying papers are entirely read and again gauged against our inclusion criteria. To ensure the objectivity of the selection, a random sample of ca. 10% of the retrieved full-texts was also handled separately by a second researcher and the results were discussed; interpretation of the selection criteria was calibrated. No conflicts were observed in the selection of key studies by the two researchers. The key studies are all included in the analysis phase of the review. From these papers, we extracted and collected: author(s); title and subtitle; year; educational context; learning analytics intervention; research objectives; operationalization of affected learning. These data are used synthesize the results of our study, as described in the next section.

## 2.4. Results

This section presents the results of our literature review. From the 1932 hits on the search terms in the seven databases, 62 key studies meet the inclusion criteria - see Figure 4. A retention of just over 3% sounds rigid, however, other literature reviews in the learning analytics domain show similar results (Bodily & Verbert, 2017; Larrabee Sønderlund et al., 2018; Mangaroska & Giannakos, 2019; Ruiz-Calleja et al., 2017). This is also in line with the earlier statement that researchers agree there are relatively few studies reporting on human-mediated interventions taking place directly in the learning process (Ferguson & Clow, 2017; Gašević et al., 2015; Larrabee Sønderlund et al., 2018; Mangaroska & Giannakos, 2019; Viberg et al., 2018).

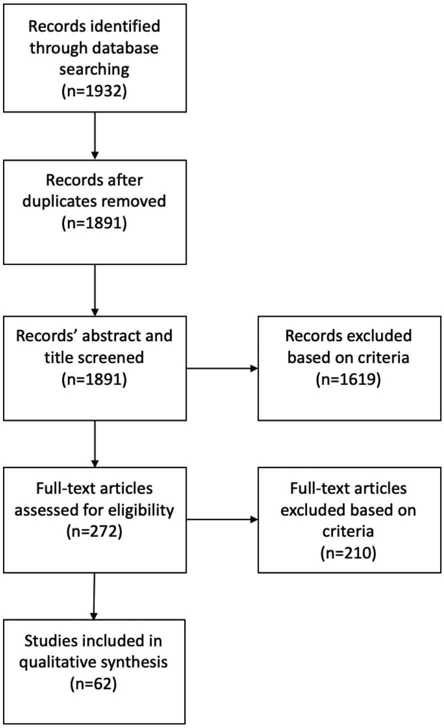


**Figure 4:** Search process results.

### 2.4.4. Descriptive view of key studies

This section provides a descriptive overview of the 62 resulting key studies. All studies with their descriptive attributes are listed in Table 4. To create this overview, we coded the key studies based on several attributes:

- *Year:* year in which the study was published;
- *Country:* country in which the study was conducted;
- *Context:* educational context in which the study was conducted;
- *Intervention Type:* category of the learning analytics intervention used in the study;
- *Data Subject:* the (role of the) person whose data was collected, analyzed and visualized in the intervention;
- *User:* the user of the intervention, i.e. the (role of the) person that had access to the learning analytics intervention and could act upon this intervention in the learning process;
- *Research Objective:* goal of the key study, classified according to the classification proposed by Papamitsiou and Economides (2014).

**Table 4:** Descriptive Overview of Key Studies

| Ref. | Authors | Year | Country | Context | Inter-vention | Data Subject | User | Research Objective |
|---|---|---|---|---|---|---|---|---|
| (Pistilli & Arnold, 2012) | Pistilli & Arnold | 2012 | US | HE | V | S | T | PP |
| (Cheng & Liao, 2012) | Cheng & Liao | 2012 | Taiwan | HE | IT | S | T | SB |
| (Smolin & Butakov, 2012) | Smolin & Butakov | 2012 | Russia | HE | O | S | SD | IA |
| (Lauría et al., 2013) | Lauría *et al.* | 2013 | US | HE | MT, PC | S | T | PP |
| (McKenzie et al., 2013) | McKenzie *et al.* | 2013 | Australia | HE | PC | S | L | PP, IR |
| (Grann & Bushway, 2014) | Grann & Bushway | 2014 | US | HE | V | S | L | IR |
| (Jayaprakash et al., 2014) | Jayaprakash *et al.* | 2014 | US | HE | IT | S | T | PP |
| (Kumar et al., 2014) | Kumar et al. | 2014 | India | HE | PC | S | L | IR |
| (Van Leeuwen et al., 2014) | van Leeuwen et al. | 2014 | Netherlands | K12 | MT | S | T | IA |
| (Yamada et al., 2014) | Yamada et al. | 2014 | Japan | HE | PC | S | L | RR |
| (Akçapinar, 2015) | Akçapınar | 2015 | Turkey | HE | O | S | L | IR |
| (Berland et al., 2015) | Berland, Davis & Smith | 2015 | US | K12 | O | S | T | SB |
| (Groba et al., 2015) | Groba et al. | 2015 | Spain | HE | V | S | T | IA |
| (Holman et al., 2015) | Holman et al. | 2015 | US | HE | IL | S | L | PP |
| (Van Leeuwen et al., 2015) | van Leeuwen et al. | 2015 | Netherlands | K12 | MT | S | T | IA |
| (Lonn et al., 2015) | Lonn et al. | 2015 | US | HE | IT, LS | S | AA | IR |
| (Marcos-García et al., 2015) | Marcos-García, Martínez-Monés & Dimitriadis | 2015 | Unknown | HE | IT | S | T | IR |
| (Martinez-Maldonado et al., 2015) | Martinez-Maldonado, Yacef & Kay | 2015 | Australia | HE | LS | S | T | IA |
| (Melero et al., 2015) | Melero et al. | 2015 | Spain | K12 | V | S | T, L | IR |
| (Nussbaumer et al., 2015) | Nussbaumer et al. | 2015 | Austria | HE | V | S | L | IR, RR |
| (Tabuenca et al., 2015) | Tabuenca et al. | 2015 | Netherlands | HE | MT | S | L | IR |

**Table 4:** Continued.

| Ref. | Authors | Year | Country | Context | Inter-vention | Data Subject | User | Research Objective |
|---|---|---|---|---|---|---|---|---|
| (Whitelock et al., 2015) | Whitelock et al. | 2015 | UK | HE | AF | S | L | IR |
| (Xiong et al., 2015) | Xiong, Wang & Beck | 2015 | Unknown | U | PC | S | L | RR |
| (Arguedas et al., 2016) | Arguedas, Daradoumis & Xhafa | 2016 | US | HE | V | S | T | PP |
| (Beheshitha et al., 2016) | Beheshitha et al. | 2016 | Canada | HE | V | S | L | IR |
| (Ben David et al., 2016) | Ben David, Segal & Gal | 2016 | Israel | K12 | PC | S | L | RR |
| (Khan & Pardo, 2016) | Khan & Pardo | 2016 | Australia | HE | V | S | L | IR |
| (Harrison et al., 2016) | Harrison et al. | 2016 | Australia | HE | IT, LS | S | U | PD |
| (Manske & Hoppe, 2016) | Manske & Hoppe | 2016 | Germany | K12 | V | S | L | IR |
| (Papoušek et al., 2016) | Papoušek, Stanislav & Pelánek | 2016 | International | All | AF | S | L | RR |
| (Sharma et al., 2016) | Sharma et al. | 2016 | Switzerland | MOOC | V | S | L | IR |
| (Siadaty et al., 2016b) | Siadaty, Gašević & Hatala | 2016 | International | W | PC | E | L | IR |
| (Siadaty et al., 2016a) | Siadaty, Gašević & Hatala | 2016 | International | W | PC | E | L | IR |
| (D. Davis et al., 2017) | Davis *et al.* | 2017 | International | MOOC | V | S | L | IR |
| (Dawson et al., 2017) | Dawson *et al.* | 2017 | Unknown | HE | MT | S | AA | PD |
| (Diana et al., 2017) | Diana *et al.* | 2017 | US | K12 | V | S | T | PP |
| (Faria et al., 2017) | Faria *et al.* | 2017 | US | K12 | IT | S | AA | PD |
| (Herodotou et al., 2017) | Herodotou *et al.* | 2017 | UK | HE | V | S | T | PD |
| (Jing & Tang, 2017) | Jing & Tang | 2017 | China | MOOC | PC | S | L | RR |
| (van Klaveren et al., 2017) | van Klaveren, Vonk & Cornelisz | 2017 | Netherlands | K12 | PC, AF | S | L | RR |
| (Perikos et al., 2017) | Perikos, Grivokostopoulou & Hatzilygeroudis | 2017 | Greece | HE | AF | S | L | IR |
| (Rosmansyah et al., 2018) | Rosmansyah, Kartikasari & Wuryandari | 2017 | Indonesia | K12 | V | S | T | IR |

**Table 4:** Continued.

| Ref. | Authors | Year | Country | Context | Inter-vention | Data Subject | User | Research Objective |
|---|---|---|---|---|---|---|---|---|
| (Seanosky et al., 2017) | Seanosky *et al.* | 2017 | Unknown | HE | V | S | L | IR |
| (Shimada & Konomi, 2017) | Shimada & Konomi | 2017 | Japan | HE | IT | S | T | SB |
| (B. Chen et al., 2018) | Chen *et al.* | 2018 | US | HE | V | S | L | IR |
| (Ciolacu et al., 2018) | Ciolacu *et al.* | 2018 | Germany | HE | MT | S | T | PP |
| (Gong et al., 2018) | Gong, Liu & Zhao | 2018 | China | HE | MT | S | T | IR |
| (Guillot et al., 2018) | Guillot *et al.* | 2018 | Canada | HE | V, IL | S | L | IR |
| (Klock et al., 2018) | Klock *et al.* | 2018 | Unknown | HE | V, IL | S | L | IR |
| (Mangaroska et al., 2018) | Mangaroska *et al.* | 2018 | Switzerland | HE | IL | S | L | IR |
| (Michos & Hernández-Leo, 2018) | Michos & Hernandez-Leo | 2018 | Spain | HE, MOOC, W | V, D | T(L) | T(L) | IR |
| (Aljohani et al., 2019) | Aljohani *et al.* | 2019 | Saudi Arabia | HE | V | S | T, S | IR |
| (Aslan et al., 2019) | Aslan *et al.* | 2019 | Turkey | K12 | V | S | T | IR |
| (Gong & Liu, 2019) | Gong &Liu | 2019 | China | HE | AF | S | T, L | IA |
| (Hubalovsky et al., 2019) | Hubalovsky, Hubalovska & Musilek | 2019 | Czech Republic | K12 | PC | S | L | RR |
| (Jeon & Song, 2019) | Jeon & Song | 2019 | Korea | K12 | IT | S | T | IA |
| (Jovanović et al., 2019) | Jovanović *et al.* | 2019 | Unknown | HE | V | S | L | IR |
| (van Leeuwen et al., 2019) | van Leeuwen, Rummel & van Gog | 2019 | Netherlands | K12 | D | SS | T | SB |
| (Lim et al., 2019) | Lim *et al.* | 2019 | Australia | HE | MT | S | T | IA |
| (Swidan et al., 2019) | Swidan *et al.* | 2019 | Israel | K12 | D | SS | T | SB |
| (Syed et al., 2019) | Syed *et al.* | 2019 | US | HE | LS | S | AA, T | SB |
| (Vargas et al., 2019) | Vargas *et al.* | 2019 | Chile | HE | IT | S | T | SB |

Since - to the best of our knowledge - no general accepted classification method is available for learning analytics interventions, we synthesized the categories for *Intervention Type* through open coding. Here, we adopted the proposed distinction between dashboards (consisting of visualization(s) of multiple indicators) and visualizations (based on a single indicator) (Schwendimann et al., 2017). Whenever information was presented to learners in a different form than visualization (e.g. plain numerical information), we coded this intervention as *Information for Learner*. Three intervention types may seem somewhat similar: *Information for Teacher (IT)*, *Message from Teacher to Learner (MT)*, and *Learner Support (LS)*. The difference is, that in type IT the intervention is passive, in the sense that information is presented to teachers (e.g. exercise completion rates for all students), whereas in intervention type MT, the teacher is supported by the intervention to actively reach out to students, e.g. to notify students at risk. In intervention type LS, the learning analytics intervention helped inform other ways of learner support, e.g. through academic advisors.

Firstly, we analyzed the educational context of the key studies. More than half of the studies describe learning analytics interventions in higher education (38), followed by K12 education (15) with only a handful of other educational contexts; see Figure 5.

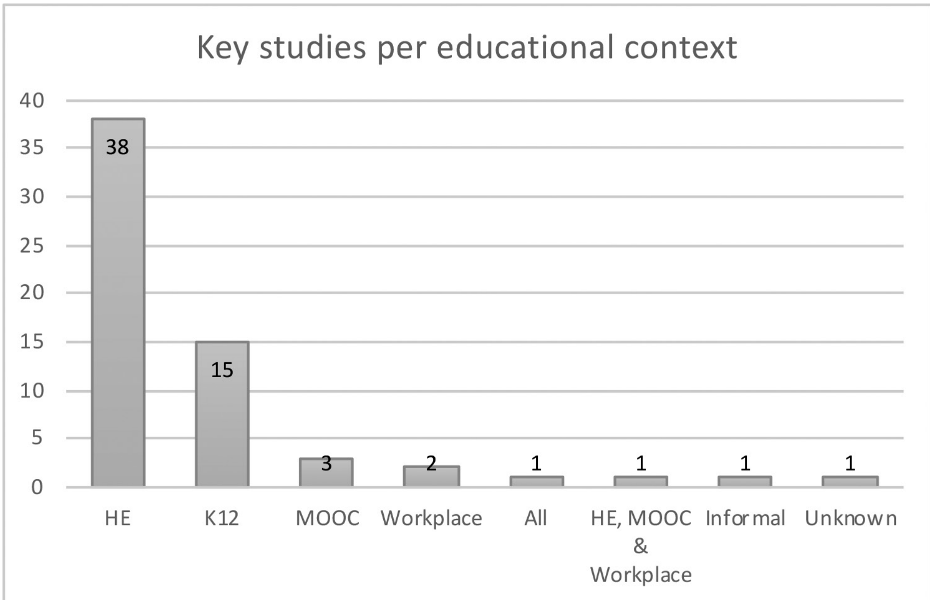


**Figure 5:** Number of key studies per educational context.

Secondly, we analyzed the number of studies per year; see Figure 6. We see that since 2015, the number of studies meeting our inclusion criteria has been approximately 10 studies per year. In the first five months of 2019, this number is already 11, which gives hope of an even larger number of this type of study in the whole of 2019. We can conclude that the number of studies in which the effects on learning of learning analytics efforts are being analyzed empirically and quantitatively has increased since 2012.

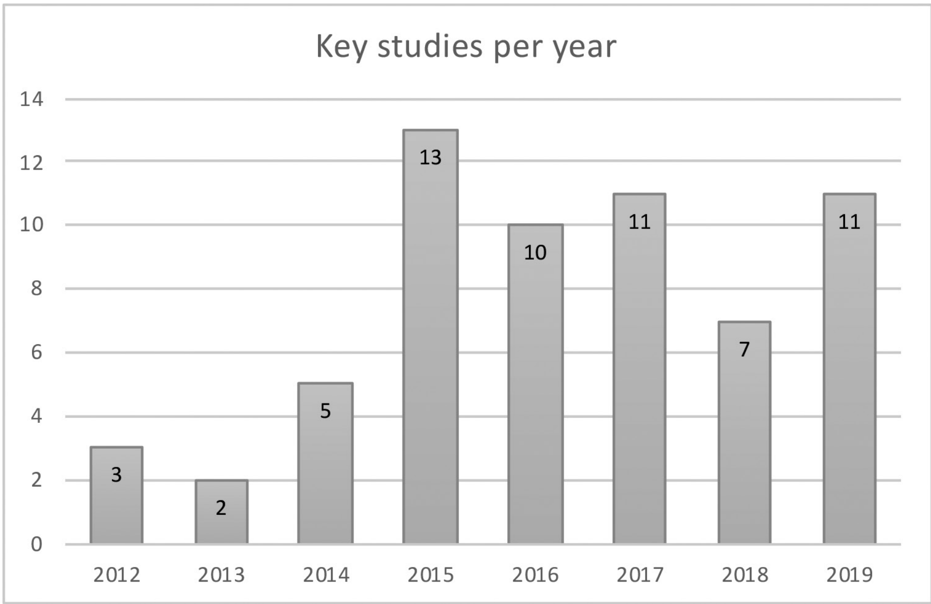


**Figure 6:** Number of key studies per year.

### 2.4.5. Classifying key studies based on affected learning

Using the preliminary classification scheme introduced in section 2.2.3., we now classify the key studies based on the different categories of affected learning, i.e., Learning environment, Learning process, and Learning outcome. To this aim, we searched the text of the key studies for description of what was measured exactly in order to analyze the effects on learning of the intervention discussed in that particular study.

We find that 53 out of 62 studies describe operational definitions that fit into a single category of our classification scheme. In Table 5, we give an overview of these single-category studies per category per year. We observe that *Learning outcome* is by far the largest category, followed by *Learning process*. *Learning environment* is the smallest category with only seven single-category key studies.

| | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | **TOTAL** |
|---|---|---|---|---|---|---|---|---|---|
| Learning environment | 1 | | 1 | 3 | | | | 2 | **7** |
| Learning process | | | | 5 | 4 | | 2 | 1 | **12** |
| Learning outcome | 2 | 2 | 4 | 4 | 5 | 9 | 3 | 5 | **34** |
| **Total** | **3** | **2** | **5** | **12** | **9** | **9** | **5** | **8** | **53** |

**Table 5:** Single-category key studies per category per year

Nine key studies contain operational definitions relating to more than one category. Table 6 gives an overview of the number of cross-categorical key studies per year. The most occurring combination of categories is Learning process & Learning outcome with eight key studies, while the number of studies in this category combination is also increasing over the years. We observe that not a single key study includes operational definitions of affected learning in all three categories.

| | 2015 | 2016 | 2017 | 2018 | 2019 | Total |
|---|---|---|---|---|---|---|
| Learning environment & learning process | | | 1 | | | **1** |
| Learning process & Learning outcome | 1 | 1 | 1 | 2 | 3 | **8** |
| **Total** | **1** | **1** | **2** | **2** | **3** | **9** |

**Table 6:** Cross-Category Studies Per Category Set Per Year

### 2.4.6. Classification in relation to research objective

We also quantitatively investigated the relation between the classification of key studies and the respective research objectives according to the classification of Papamitsiou & Economides (2014); see Table 7. Note that in this table, we included all combinations of classifications and research objectives to get a complete overview, i.e. if a key study relates to more than one category from our classification scheme or has more than one research objective, we counted all combinations. At first glance, the results do not look surprising; for example, key studies that measure affected learning in the category *Learning environment* mostly aim to improve assessment & feedback services, while studies that aim to predict performance usually operationalize affected learning in the category *Learning outcome*. However, it might actually not be that straightforward. Consider those studies in which the research objective is Prediction of Performance: in only two out of nine key studies, the learner is the user of the intervention (Holman et al., 2015; McKenzie et al., 2013), while in the other seven studies, the learning analytics intervention has teachers as target users. Equipping teachers with a learning analytics tool might be an indication that we are in fact trying to intervene on aspects of learning in the category *Learning environment*, with the ultimate goal of increasing *Learning outcome*. In this chain of reasoning, we might need to consider measuring both the intermediate and the ultimate effects of our interventions by incorporating operational definitions from multiple categories in our studies.



| | Improve Assessment & Feedback Services | Increase (self) Reflection & (self) Awareness | Prediction of Dropout and Retention | Prediction of Performance | Recommendation of Resources | Student Behavior Modeling |
|---|---|---|---|---|---|---|
| Learning environment | 5 | | | | | 3 |
| Learning process | 2 | 15 | | 1 | 3 | 1 |
| Learning outcome | 3 | 17 | 4 | 7 | 9 | 4 |
| **Total** | **10** | **32** | **4** | **8** | **12** | **8** |

**Table 7:** Classification of Key Studies Related to Research Objective

### 2.4.7. Analysis and refinement of the classification scheme

In this section, we analyze the retrieved operational definitions for each of the three categories of the preliminary classification scheme. We distilled subcategories through iterative open coding until convergence occurred, which we also link to relevant literature.

#### *2.4.7.1. Learning environment*

Although the optimization of the learning environment is explicitly mentioned in the commonly accepted definition of learning analytics (LAK, 2011), with only eight key studies this category is the smallest within our research. We found nine different operational definitions in this category, out of which we distilled three subcategories: *Teacher awareness*, *Teacher productivity*, and *Learning materials}*.

*Teacher awareness* relates to operational definitions such as detection, attention and interaction by teachers (Martinez-Maldonado et al., 2015; Shimada & Konomi, 2017; van Leeuwen et al., 2019; Van Leeuwen et al., 2014, 2015). Schwendimann et al. also mention 'teacher awareness (of students)' as a construct for the evaluation of learning analytics dashboard (Schwendimann et al., 2017). Teacher productivity relates both to efficiency and effectiveness: operational definitions in this subcategory include e.g. the number of messages a teacher sends (Van Leeuwen et al., 2014, 2015), the time it takes a teacher to respond or assess time (Groba et al., 2015; van Leeuwen et al., 2019), and the quality of assessment by a teacher (Swidan et al., 2019; van Leeuwen et al., 2019). We also recognize this subcategory in a previous review, that identifies 'productivity and effectiveness in teaching' as an outcome of learning analytics interventions (Wong & Li, 2020). A single key study does not fit into the above-mentioned subcategories: Smolin & Butakov (2012) used an operational definition related to the quality/suitability of learning materials. We therefore also include the subcategory *Learning materials* in our refined classification scheme. Considering the common learning analytics research goal 'recommendation of resources' (Papamitsiou & Economides, 2014), we may expect more operational definitions in this subcategory in further research.

### *2.4.7.2. Learning process*

The learning process relates to learning activity-focused activities. We found a total of 21 key studies that measured to what extent different aspects of the learning process were affected by learning analytics interventions. We found 14 different operational definitions, from which we distilled five subcategories: *Learner awareness, Learner productivity, Self-regulated learning, Engagement*, and *Online activity & behavior.*

The first two subcategories are similar to the first two subcategories in *Learning environment*; here the focus is on the learner instead of the teacher. Examples of operational definitions for *Learner awareness* are plagiarized post ratios (Akçapinar, 2015) and making predictions about grades by students (Holman et al., 2015). Examples of operational definitions for Learner productivity are study time, practice time, number of exercises made (Tabuenca et al., 2015; van Klaveren et al., 2017) and time spent on solving questions (Khan & Pardo, 2016). We recognize these two subcategories in the earlier discussed reviews: 'awareness of students by [..]other peers' (Schwendimann et al., 2017), and 'enhanced productivity/effectiveness in learning' (Wong & Li, 2020).

We also found three key studies using operational definitions related to *Self-regulated learning* (SRL). SRL has three important characteristics: a) self-observation, b) self-judgement and c) self-reactions (Zimmerman, 1989). Operational definitions in this subcategory include pre- and post-questionnaire scores on self-assessment and the application of SRL (Melero et al., 2015; Siadaty et al., 2016b) and the use of metacognitive tools (Nussbaumer et al., 2015). SRL skills also can be the intended learning outcome of a learning process (see the discussion of the category Learning outcome in the following section. In this subcategory, the focus is on (evidence for) the application of SRL in the learning process. Measuring SRL is not straightforward; it is also argued that the measurement of SRL is intertwined with the intervention based on this measurement (Panadero et al., 2016).



*Engagement* of learners is increasingly used as a measure of success of educational institutions (Tai et al., 2019). Noticeable is the fact that in the 3P model, an affective learning outcome is involvement, which has a strong relation to engagement and the learning process. We decide to make *Engagement* a subcategory of *Learning process* - in which we follow (Joksimović et al., 2018)- however, we also recognize an ongoing discussion in the field on engagement, how to model, operationalize and measure this construct (Reschly & Christenson, 2012; Tai et al., 2019). Examples of operational definitions we found in our review that relate to engagement are social interactions (B. Chen et al., 2018; Marcos-García et al., 2015; Michos & Hernández-Leo, 2018) and emotional changes (Aslan et al., 2019). We believe this is only a limited view of the complex construct of engagement; many other operational definitions of engagement are available. Joksimović et al. provide metrics for e.g. academic, behavioral, cognitive, and affective engagement (Joksimović et al., 2018).

Finally, we decided to separate the two subcategories *Engagement* and *Online activity & behavior*, even though the latter is often used as a proxy for the former. We found a set of operational definitions in Online activity & behavior such as event count (Siadaty et al., 2016b), frequency of accessing the LMS (Aslan et al., 2019; Swidan et al., 2019), quantity and/or quality of discussion board posts (Aljohani et al., 2019; Beheshitha et al., 2016; Gong et al., 2018; Klock et al., 2018; Mangaroska et al., 2018; Michos & Hernández-Leo, 2018), and the use of note-taking functionality (Shimada & Konomi, 2017). Measures of activity in a VLE or an LMS are also mentioned in many recent reviews as operational definitions of affected learning (Larrabee Sønderlund et al., 2018; Schwendimann et al., 2017; Viberg et al., 2018; Wong & Li, 2020). Engagement can be measured in much more diverse ways than simple event counts, and we hope to emphasize this by separating these two subcategories in our refined classification scheme.

### *2.4.7.3. Learning outcome*

Containing 42 key studies, this last category is by far the largest in our research. We found 14 different operational definitions in this category, which we grouped into three subcategories: *Knowledge and skills, Learning gain*, and *Retention and dropout*. The first two subcategories are more focused on individual learners (at the course level), whereas the third relates to larger groups of learners (at the department level). In many of the key studies, we found concepts such as academic performance, academic achievement and academic success. We argue that these concepts are too abstract for transparent evaluation of learning analytics interventions (and for learning technologies in general), hence we used more explicitly named subcategories.

A learner can demonstrate the acquisition of *Knowledge and skills* as a product of their learning process. Most key studies in this category operationalize affected learning through grades or test scores or scores (Ben David et al., 2016; Cheng & Liao, 2012; D. Davis et al., 2017; Diana et al., 2017; Gong et al., 2018; Grann & Bushway, 2014; Guillot et al., 2018; Jayaprakash et al., 2014; Khan & Pardo, 2016; Klock et al., 2018; Kumar et al., 2014; Papoušek et al., 2016; Perikos et al., 2017; Rosmansyah et al., 2018; Seanosky et al., 2017; van Klaveren et al., 2017; Whitelock et al., 2015; Xiong et al., 2015; Yamada et al., 2014) which is a direct assessment of learning as performance on a task (e.g. an exam or final test) (Rome, 2011). Although grades may seem to be a direct operational definition, this is debatable. Grades can be regarded as a proxy for learning, as they often comprise a combination of learning outcomes or include non-related corrections like extra credits for certain activities (Teaching Commons, 2019). Other operational definitions might capture knowledge and skills more directly, such as the quality of an artifact created by the learner (Berland et al., 2015; Mangaroska et al., 2018). Remarkably, some of the key studies claim to affect aspects which one would expect in one of the other categories - e.g., saving time for teachers in monitoring the progress of the learning process of students (Rosmansyah et al., 2018) but the operational definitions actually fall in the *Learning outcome category* (e.g., grades or scores). That is, the product or outcome of the learning process is measured rather than the actions performed during this learning process or in the learning environment. Moreover, we observe that in some studies, researchers wish to improve higher-order learning outcomes, such as self-regulated learning skills (D. Davis et al., 2017; McKenzie et al., 2013). Since these higher-order skills are meta-cognitive and difficult - if not impossible - to measure, these researchers presumably chose to measure the effects in grades or test scores instead.

*Knowledge and skills* and the second subcategory *Learning gain* are closely related; we separated them because the former relates to absolute operational definitions (such as grades) and the latter relates to relative operational definitions (such as the difference between pre- and a post-tests (Lonn et al., 2015; Sharma et al., 2016)), emphasizing the difference in learning a learner has achieved. The concept of *Learning gain* captures the idea that learning is visible through a change over time in a learner's behavior, attitude and/or knowledge. There is no standard definition, conceptualization or measurement (instrument) to assess learning gain; a conceptual framework with a set of measurement tools is currently being developed for English higher educational institutes (Vermunt et al., 2018). In the conceptual framework proposed in Vermunt et al. (2018), a distinction is made between four components (cognitive, metacognitive, affective and socio-communicative) and three cross-cutting dimensions (view of knowledge and learning, research attitude and moral reasoning).



The final subcategory is *Retention and dropout*, which relates to larger groups of learners and captures 'academic persistence' in terms of e.g. withdrawal rates and absence (Faria et al., 2017; Jayaprakash et al., 2014; Lauría et al., 2013; Pistilli & Arnold, 2012), student retention (Dawson et al., 2017; Herodotou et al., 2017) and reregistration rates (Grann & Bushway, 2014). Long & Siemens (2011) distinguish between learning analytics at course level and departmental level. Departmental variables may consider a more long-term effect of learning analytics, which has been posed as an important feature of future learning analytics research (Gašević et al., 2015).

The above synthesis leads us to the refined classification scheme of Figure 7. We use this refined classification scheme to give an overview of all operational definitions identified in the key studies of this review; see Table 8.

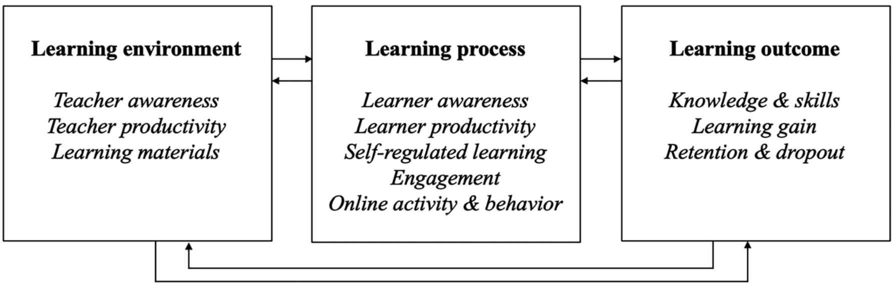


**Figure 7:** Refined classification scheme for operational definitions of learning affected by learning analytics interventions.

| **Category** | Subcategory / construct | Operational definition | Key studies |
|---|---|---|---|
| **Learning environment** | Teacher awareness | Detection of problematic student groups | (Van Leeuwen et al., 2014, 2015) |
| | | Teacher attention, teacher interaction | (Martinez-Maldonado et al., 2015) |
| | | Synchronization between teacher and student | (Shimada & Konomi, 2017) |
| | Teacher productivity | Number of messages sent by teacher | (Van Leeuwen et al., 2014, 2015) |
| | | Time it takes a teacher to assess a student | (Groba et al., 2015) |
| | | Percentage of correct answers (identification) regarding group work | (Swidan et al., 2019) |
| | | Response time | (van Leeuwen et al., 2019) |
| | | Number of correctly detected problem types | (van Leeuwen et al., 2019) |
| | Learning materials | Validity, reliability of exam | (Smolin & Butakov, 2012) |

| Category | Subcategory / construct | Operational definition | Key studies |
|---|---|---|---|
| Learning process | Learner awareness | Plagiarized post ratios | (Akçapinar, 2015) |
| | | Making predictions about grades by students | (Holman et al., 2015) |
| | | Revision of artifact made | (Manske & Hoppe, 2016) |
| | Learner productivity | Study time, practice time, number of exercises made | (van Klaveren et al., 2017), (Tabuenca et al., 2015) |
| | | Time spent on solving questions, higher level of difficulty of questions | (Ben David et al., 2016) |
| | Self-regulated learning | Self-assessment pre- & post-questionnaire scores<br>Self-regulated learning pre- & post-questionnaire scores | (Melero et al., 2015), (Siadaty et al., 2016b) |
| | | Use of metacognitive tools, application of SRL cycle (plan, learn, assess, reflect) | (Nussbaumer et al., 2015) |
| | Engagement | SNA indicators, social interactions and engagement | (B. Chen et al., 2018), (Marcos-García et al., 2015), (Michos & Hernández-Leo, 2018) |
| | | Emotional duration changes | (Aslan et al., 2019) |
| | | Counts of events, centrality measures | (Siadaty et al., 2016a) |
| | | Number of posts, threads, messages or discourse features | (Beheshitha et al., 2016) (Gong et al., 2018)(Klock et al., 2018) (Michos & Hernández-Leo, 2018) (Aljohani et al., 2019) |
| | Online activity & behavior | Utilization ratio of note-taking functionality | (Shimada & Konomi, 2017) |
| | | Online learning behaviors (Discussion posts, Resources browsing, Self-evaluation, Peer evaluation, Tasks submission) | (Gong & Liu, 2019) |
| | | Online activity, frequency of accessing LMS, frequency of accessing discussion board | (Aljohani et al., 2019; Lim et al., 2019) |

| **Category** | Subcategory / construct | Operational definition | Key studies |
|---|---|---|---|
| **Learning outcome** | Knowledge and skills | Grades | (D. Davis et al., 2017; Diana et al., 2017; Gong & Liu, 2019; Grann & Bushway, 2014; Guillot et al., 2018; Jayaprakash et al., 2014; Khan & Pardo, 2016; Klock et al., 2018; Kumar et al., 2014; Lauría et al., 2013; Lim et al., 2019; McKenzie et al., 2013; Pistilli & Arnold, 2012; Rosmansyah et al., 2018; Seanosky et al., 2017; Syed et al., 2019; Tabuenca et al., 2015; Vargas et al., 2019; Whitelock et al., 2015) |
| | | Answers to reference questions | (Papoušek et al., 2016) |
| | | Depth, rarity, quality, and specificity of program | (Berland et al., 2015) |
| | | Number of unit tests passed | (Mangaroska et al., 2018) |
| | | Quiz scores, test scores, final exam scores | (Ben David et al., 2016; Cheng & Liao, 2012; Gong et al., 2018; Jovanović et al., 2019; Lim et al., 2019; McKenzie et al., 2013; van Klaveren et al., 2017; Xiong et al., 2015; Yamada et al., 2014) |
| | | Score of the game | (Arguedas et al., 2016) |
| | | Exercise performance (completion time & success rate) | Hubalovsky, Hubalovska & Musilek (Hubalovsky et al., 2019) |
| | | Course completion or failure rates | (Ciolacu et al., 2018; D. Davis et al., 2017; Faria et al., 2017) |
| | Learning gain | Difference between pre- and post-test | (Aslan et al., 2019; Jeon & Song, 2019; Perikos et al., 2017; Sharma et al., 2016) |
| | | Mastery scores (pre- and post-) | (Lonn et al., 2015) |
| | Retention and dropout | Withdrawal rates, absence | (Faria et al., 2017; Jayaprakash et al., 2014; Lauría et al., 2013; Pistilli & Arnold, 2012) |
| | | Student retention | (Dawson et al., 2017; Herodotou et al., 2017) |
| | | Reregistration rate | (Grann & Bushway, 2014) |
| | | (Revenue from) student enrollment | (Harrison et al., 2016; Jing & Tang, 2017) |

**Table 8:** Operational Definitions of Affected Learning Positioned Within our Refined Classification Scheme

## 2.5. Discussion



The aim of this study was to provide an answer to the research question: *In what way does existing literature on learning analytics interventions operationalize affected learning?* The first conclusion is that, from 1932 search hits on learning analytics, only 62 describe quantitative, measurable effects of complete learning analytics cycles in authentic learning context. This is a noticeable shortcoming and in line with previous research that concluded that not enough studies make a connection to the next stage of the learning analytics cycle. i.e., `not enough published work is making clear how the move will be made from researching the data to optimizing the learning' (Ferguson & Clow, 2017). As we gathered evidence from a wide range of scientific databases, our study thoroughly underpins their conclusion, which was based on only those studies that were included in the LACE Evidence Hub. We concur with the conclusions of several other reviews in the field that the number of studies providing evidence for the (positive) impact of learning analytics on learning currently is low (Schwendimann et al., 2017; Viberg et al., 2018; Wong & Li, 2020).

By analyzing these 62 key studies, we identified different operational definitions of learning which can be affected with learning analytics interventions. The operational definitions are positioned according to a classification scheme with three categories and eleven subcategories: *Learning environment, Learning process*, and *Learning outcome*. Our study facilitates improved positioning of empirical research on learning analytics interventions based on concrete operational definitions, which in turn helps to better compare and generalize studies. We hope to advance the field in this respect, motivated by recent calls for (the use of) a standard set of constructs evaluating the impact of learning analytics studies (Larrabee Sønderlund et al., 2018; Mangaroska & Giannakos, 2019; Schwendimann et al., 2017). Our classification scheme with suggestions for operationalization could be used in conjunction with a framework for systematic development, implementation and evaluation of learning analytics interventions (Rienties et al., 2017). Our results can also be used in the evaluation of learning technologies in general, since the impact on learning can be measured similarly for other technologies.

This systematic literature review shows that key studies mostly relate to the two subcategories *Learning process - Online activity & behavior* and *Learning outcome - Knowledge and skills.* This is not surprising, since grades, test scores and LMS log data are easily gathered. Merely nine key studies report on operational definitions in more than one category, even though cross-categorical learning analytics provide a better, multi-perspective view on learning. The need for multidimensional metrics for learning is supported by Joksimović et al. (2018). Moreover, given the complex nature of applying learning analytics interventions in practice, measuring the effect of learning analytics interventions on a wider spectrum of aspects can give more insight into their workings on different actors and processes involved. We believe this is not exclusively important for learning analytics research, but is crucial in the development of learning technologies in general because of the shared goal of optimizing learning through technological interventions. We observe that all cross-categorical studies have appeared in the most recent years (since 2014), which might be an indication that the need for this type of study is increasingly acknowledged in the learning analytics field.

### 2.5.1. Recommendations

In order to justify the use of data analytics within educational processes, the effects of learning analytics interventions on learning must be clear and well-defined. In a recent book chapter, Wise (2019) describes the various pedagogical uses for which learning analytics are used. All of them focus on improving either the learning process or the learning environment. It makes sense to empirically evaluate whether learning analytics efforts indeed have done so, by measuring effects of learning analytics interventions on these particular aspects directly.

Finding the operational definition(s) of affected learning in a research paper was not always straightforward; these operational definitions could be found in the methodology section, the analysis or the result section. Often, the dependent variable of a study would be named as an abstract construct in most of the paper (e.g. academic achievement or engagement), whereas the concrete operational definition or measurement instrument would only be mentioned explicitly once in the result section, without justification on why this specific operationalization was adopted. We share this observation with Joksimović et al. (2018), who state that a lack of specificity on used concepts and measures posed a significant challenge for their study. We recommend researchers to be clear and transparent throughout the paper on which operational definitions are used to measure learning affected by learning analytics interventions.

Some of the papers we encountered during this study do report on potential improvements gained by learning analytics interventions but do not quantify the actual effects by operationalizing and measuring affected learning. This is in line with the observations from the review by Viberg et al. (2018). By describing those effects, more evidence about the benefits of learning analytics on education can be gathered, consequently strengthening the field in general. We suggest the use of our research outcomes for reporting on and comparing learning analytics results in both research and practice. It is a first step to be clear and transparent about the operational definitions and measurement instruments we use in our empirical evaluations, before the learning analytics field as a whole can standardize these operational definitions in order to ultimately compare effect sizes in the same way this is done in research fields with a longer tradition such as medicine and psychology.

As mentioned before, grades can be regarded as a proxy for learning. Recently, Guillot et al. (2018) also concluded that grades (alone) are no suitable way to operationalize the impact of learning analytics systems. The problem that operationalizing affected learning results in shallow proxies for learning extends well beyond grades, since the data that is available to the researchers often limits which measurements can be used; this need not be a huge problem, as long as researchers are transparent on which operational definitions are used. Rienties, Toetenel and Bryan give a good example of such transparency: in their study they state 'LMS activity should only be regarded as a proxy for student engagement in formal online activities, as at this point in time the OU does not systematically collect data about formal or informal offline activities' (Rienties et al., 2015). We emphasized this perspective by separating the subcategories *Engagement* and *Online activity & behavior.*

Moreover, we observe that higher-order learning outcomes, such as self-regulated learning skills, are difficult - if not impossible - to operationalize and measure. Further research could explore alternative operational definitions that fit the higher-order nature better than grades or test scores do.

Gašević et al. (2015) urge us to remember that "learning analytics are about learning". In line with this statement, and based on the outcomes of this study, we recommend learning analytics researchers and educational institutes to move away from mere performance-based evaluation of learning analytics projects and include measurements related to learning processes and learning environment as well, as that is also a core objective of learning analytics (Long & Siemens, 2011). Regardless of the dominant learning theory within an institute, a more complete view on learning is taken by adopting a multi-perspective operationalization from more than one category of our classification scheme.

### 2.5.2. Limitations

We used a classification scheme based on the 3P model - in line with the approach of Joksimović et al. (2018) - to categorize the operational definitions we found. Other approaches to classify could lead to different insights, since a choice for a specific classification scheme introduces a level of subjectivity. In relation to our category Learning outcome, Viberg et al. (2018) use similar, but slightly different categories in their review of learning analytics evidence: knowledge acquisition, skill development and cognitive gain. Bart Rienties et al. (2017) propose to evaluate the impact of learning analytics interventions using the Attitude, Behavior and Cognition (ABC) model. Attitude and behavior have the strongest relation with our category Learning process and cognition with Learning outcome. However, although cognition is often measured through summative assessments, operational definitions also can incorporate more formative learning activities, such as discussion forum activity or blog postings (Rienties et al., 2017).

Our goal was to systematically review in what way literature on learning analytics interventions operationalizes affected learning. In order to do so, we only included empirical, quantitative results from the evaluation of learning analytics interventions in our study in the same way Joksimivić et al. did in their study on modeling learning in MOOC research (Joksimović et al., 2018). However, several studies use tools, techniques or methods as an intervention, even though they do not rely on data analytics itself. These papers then use data (analytics) to describe the effect the intervention has on learning. Although this provides insight in the variables used to measure affected learning, these studies were disregarded as they do not meet our inclusion criterion demanding interventions based on learning analytics, which is an important step within the learning analytics cycle and the focus of this study. Furthermore, qualitative studies will probably yield complementary, rich insights; we believe such studies are worthy of a separate literature review. Future research might adopt broader inclusion criteria and extend the current findings with a larger set of key studies, thereby enhancing our results and identifying more and different operational definitions of affected learning.

# PART II

## Rationale

# CHAPTER

# 3

# 3. A First Step Towards Learning Analytics: Implementing an Experimental Learning Analytics Tool[2]

*The educational domain is momentarily witnessing the emergence of learning analytics - a form of data analytics within educational institutes. Implementation of learning analytics tools, however, is not a trivial process. This research-in-progress focuses on the experimental implementation of a learning analytics tool in the virtual learning environment and educational processes of a case organization - a major Dutch university of applied sciences. The experiment is performed in two phases: the first phase led to insights in the dynamics associated with implementing such tool in a practical setting. The second - yet to be conducted - phase will provide insights in the use of pedagogical interventions based on learning analytics. In the first phase, several technical issues emerged, as well as the need to include more data (sources) in order to get a more complete picture of actual learning behavior. Moreover, self-selection bias is identified as a potential threat to future learning analytics endeavors when data collection and analysis requires learners to opt in.*

## 3.1. Introduction

Data analytics is already applied in many industries. The educational domain, however, has only recently started using data to improve its processes (Ferguson, 2012). Analytical activities aimed at improving education at a micro-level of educational institutes is called learning analytics: "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environment in which it occurs" (LAK, 2011). Common research objectives include modelling student behavior in virtual learning environments and prediction of performance (Papamitsiou & Economides, 2014). The learning analytics process comprises four steps: 1) learners generate data, 2) these data are captured, collected and stored, 3) analysis and visualization of the data, and 4) the design and use of data-driven pedagogical interventions (Clow, 2012). The process is a cycle, as the effects of the interventions can again be measured, analyzed, visualized et cetera – see Figure 8.

---

2 This work was originally published as: Knobbout, J.H., and van der Stappen, E.J. (2017). A First Step Towards Learning Analytics: Implementing an Experimental Learning Analytics Tool. *BLED 2017 Proceedings*. Paper 28.

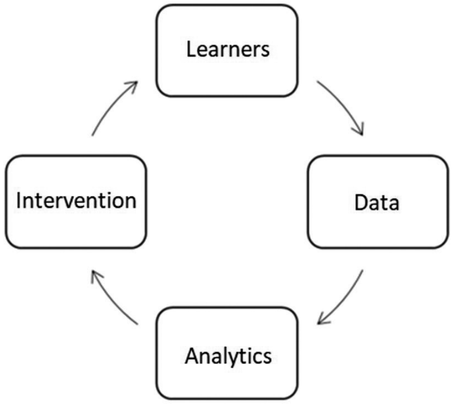


**Figure 8:** The Learning Analytics Cycle (Clow, 2012)



Many examples of small-scale learning analytics initiatives exist (Fidalgo-Blanco et al., 2015; Lonn et al., 2015; Macfadyen & Dawson, 2010; Romero-Zaldivar et al., 2012). Our case organization, however, has only very limited experience with the application of learning analytics. To overcome this lack of experience, a learning analytics experiment is conducted. This study aims at identifying and understanding some of the dynamics associated with the implementation of learning analytics tools within the case organization and in educational institutes in general. Two research questions will be answered: 1) what issues are encountered when implementing an experimental learning analytics tool in the case organization's virtual learning environment, and 2) in what way can the visualizations in the learning analytics teacher dashboard be used to design and perform timely pedagogical interventions? In this study, we will implement an externally developed learning analytics tool and research what barriers need to be overcome in order to apply the visualizations of the tool to education.

The remaining of this paper is structured as follows. First, the background of learning analytics processes and our experiment will be provided. We then elaborate on the primarily findings of the experiment conducted thus far. Finally, future work will be described.

## 3.2. Background

Higher educational institutes implementing learning analytics processes face several difficulties, e.g., changing existing information systems by implementing a learning records store (LRS) and customizing data streams (Del Blanco et al., 2013); managing the increase in workload for teachers (Whale et al., 2013); and making sure all activities are in compliance with privacy legislation (Sclater & Bailey, 2015). Greller & Drachsler (2012) provide a generic learning analytics framework with six critical dimensions to consider whilst setting up learning analytics services - see Figure 9.

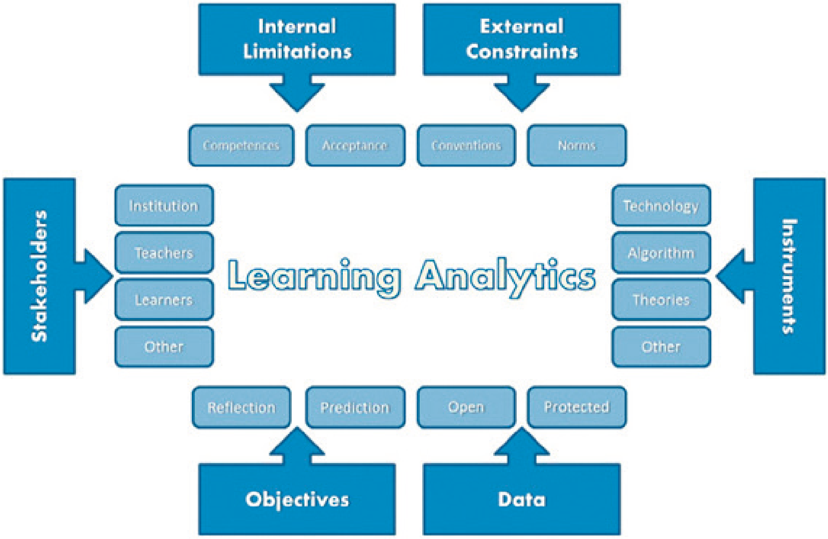


**Figure 9:** The Learning Analytics Framework (Greller & Drachsler, 2012)

In order to help Dutch higher educational institutes overcome the aforementioned difficulties, SURFnet[3] set up the learning analytics experiment (Manderveld, 2016). The experiment provides all instruments - IT-architecture, data standard, algorithms - required to answer five pre-defined pedagogic questions. These questions are sourced from previous research aimed at identifying questions relevant for Dutch educators (Berg et al., 2015). Setting up the tool is relatively easy, as the only necessarily activities involve putting tracking codes on the pages and learning materials in the virtual learning environment. Since the codes can easily be copy-pasted, only basic computer skills are required. Once placed, the codes allow data to be stored in a learning records store, from which the data can be analyzed and visualized in a dashboard. See Figure 10 for the learning analytics architecture and Figure 11 for the resulting dashboard as seen with a teachers' account. Teachers can see the anonymized activities of all learners in the experiment; learners have their individual dashboard and can view their own

3 Part of SURF - the collaborative ICT organisation for Dutch education and research.

activities. To ease compliance with Dutch privacy laws, students must opt-in to allow their data being captured and analyzed.

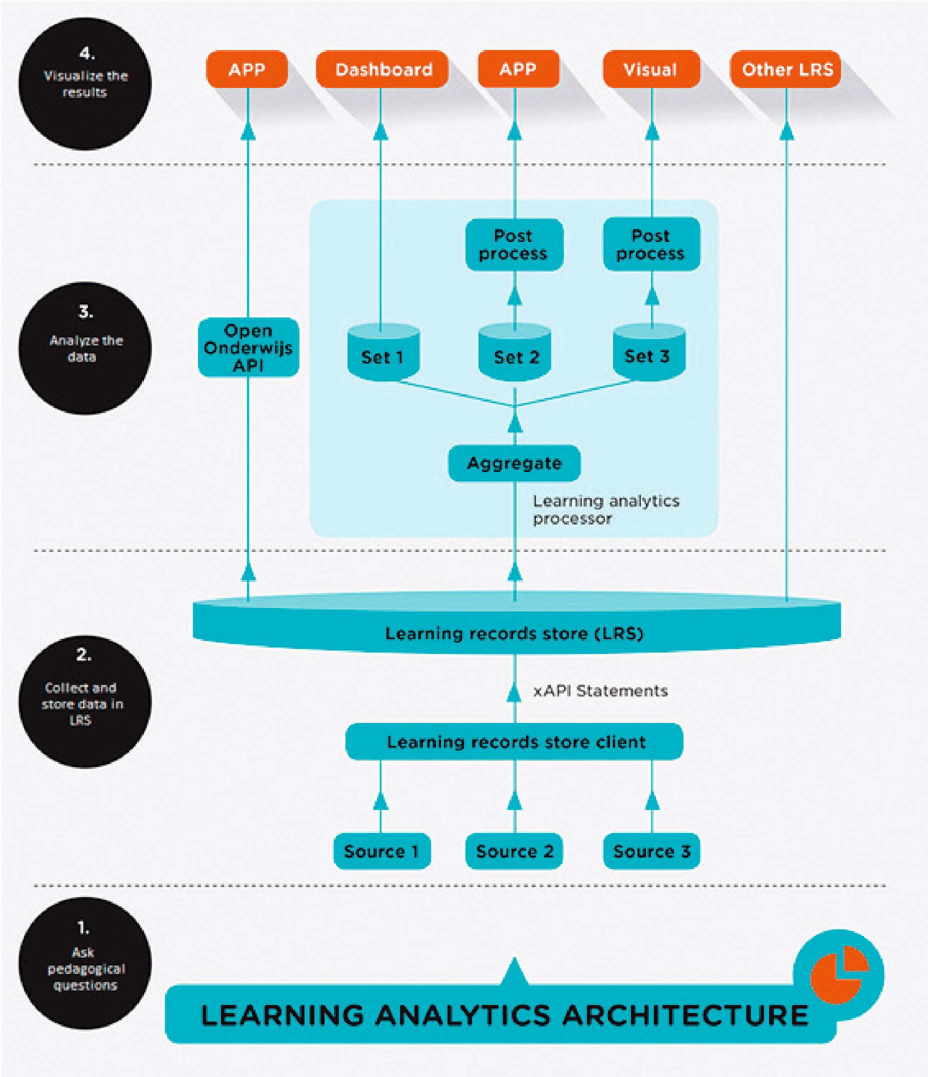


**Figure 10:** Learning Analytics Architecture (Manderveld, 2016)

3

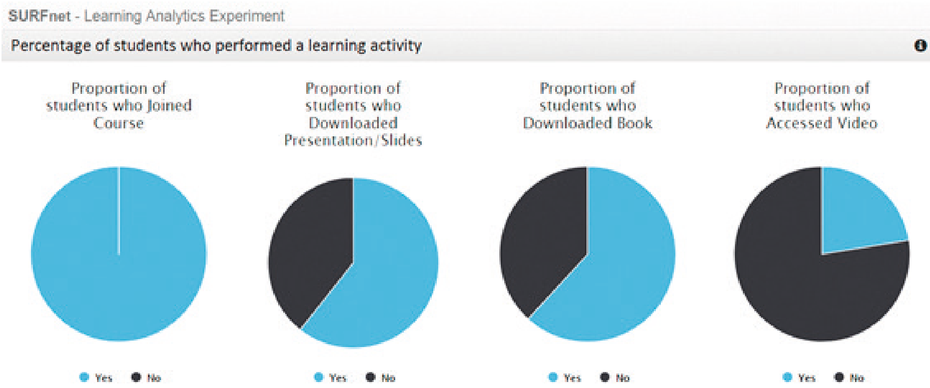


**Figure 11:** Dashboard As Seen With Teacher Account

Our case organization is a large university of applied sciences in the Netherlands. At two locations within the organization the Learning Analytics Experiment is conducted; at the Institute of Engineering and Design (IED) and at the Institute of Teacher Education (ITE). At IED, data is collected by tracking activities performed by first-year students of the Business Engineering undergraduate program enrolled in a statistics course. ITE, on the other hand, educates future secondary school English teachers - the experiment is here conducted in the courses *Curriculum Design* and *ICT-rich Education*.

## 3.3. Research objectives and method

The objective of this study is to research to what extent it is possible to perform learning analytics activities in the case organization's virtual learning environment and what barriers are encountered when doing so (first phase of the experiment), and to research in what way pedagogical interventions can be designed and performed during the course by teachers based on the visualizations in the learning analytics dashboard (second phase of the experiment). We do this by answering the following research questions:

1. What issues are encountered when implementing an experimental learning analytics tool in the case organization's virtual learning environment?
2. In what way can the visualizations in the learning analytics teacher dashboard be used to design and perform timely pedagogical interventions?

We use the framework of Greller & Drachsler (2012) to create a shareable description of context parameters for this learning analytics project. Below, we describe the first phase of the learning analytics experiment in Table 9. In the section Future Work of this paper, we will provide a description of the proposed second phase. After the first phase of the experiment, all involved teachers and researchers from SURFnet evaluated the process and outcomes during a focus group. Based on open observations, all experiences worth pointing out by the participants were discussed and notated. As all teachers involved encountered the same issues, consensus was reached fast. This led to the description of the preliminary results, as shown in the next section.

| Dimension | Value |
|---|---|
| Stakeholders | Students of two institutes of the case organization, participating in one of three courses (n=294). Three lecturers teaching the courses and actively involved in the experiment. SURFnet as external facilitator of the experiment. IT-department to allow SURFconext connection. |
| Objective | Implementing a learning analytics tool integrated in the organization's virtual learning environment. |
| Data | Activities of students in virtual learning environment, measured via the experimental tool. |
| Instruments | SURFnet's Learning Analytics Architecture, including teacher and student dashboards. Virtual learning environment of case organization. |
| External limitations | Privacy laws call students to opt-in for the experiment. Connection between case organization's and SURFnet's architecture. |
| Internal limitations | Time necessary to place tracking code on target pages and learning materials in the virtual learning environment. |

**Table 9:** Learning Analytics Experiment Case Description

## 3.4. Preliminary results

The first round of the experiment resulted in several experiences and insights. In this section, we will elaborate on the most pressing results. First, an overview of encountered issues is provided. These are then categorized according to the dimensions of the learning analytics framework (Greller & Drachsler, 2012).

### 3.4.1. Technical issues

Although it is relatively easy to set up the tracking at the virtual learning environment, some technical issues arose. During the experiment, it became clear that not all activities were tracked. Root cause for this anomaly were the necessity to opt-in for the experiment; students who opted-in only provided permission to capture data from the specific device they used to subscribe. This results in incomplete datasets and renders extensive data analysis useless. The causing issue has been resolved, so this problem will not appear in the second round of the experiment.

Another encountered technical issue relates to the SURFconext connection between the case organization's virtual learning environment and the SURFnet learning analytics dashboard. With SURFconext, it is possible for students to log in the learning analytics dashboard by using their own institutional username and password. This saves the need to create a new account. To establish the connection, the institutional SURFconext contact person must login his administrator account and provide permission to do so. At the beginning of the experiment, however, the connection was not allowed until only a few hours before the start of the courses. This almost led to cancellation of the experiment as the researchers did not want to confuse students with non-working dashboards and connections. This experience shows the need for institutional-broad support and cooperation in order to make learning analytics work.

### 3.4.2. Self-Selection bias

Students need to opt-in for the experiment, allowing their data being captured and analyzed. Of the 234 Business Engineering students who enrolled in the course (both first-year students and students from later years wanting to retake the exam), 89 opted-in - only 38%. Similar proportions were found at the other institute. After the final exam of the course, the difference between the final grade of the students who participated in the experiment and those who did not were calculated[4] using SPSS. Participants scored an average grade of 5.1 (*sd = 2.2*) and non-participants an average grade of 4.3 (*sd = 2.3*). An independent t-test shows there is a significant difference ($p = 0.032$) between the two groups and students who participated in the experiment scored on average better than those who did not. As no interventions were initiated based on the information in the dashboard, it is suspected that self-selection takes place. Future initiatives involving voluntary participation must account for this effect.

4 Students who failed to show up at the exam were excluded from the analysis as they do not get a grade but are marked with NA - "Not Available".

### 3.4.3. Need for more data

In order to perform effective interventions, rich data is required (Conijn, Snijders, Kleingeld, et al., 2016; Tempelaar et al., 2015). In our experience, the current data is too poor to use for interventions. The lack of sufficient data can partly be traced back to the teachers who put too few tracking codes on their course page of the virtual learning environment. As one of the involved teachers puts it: ``next time I will track everything: every page, every article and every video." The experiment also relied on data from the virtual learning environment only. This provides only one side of the story - anecdotal observations of students using the online materials showed that they sometimes jointly sit together at a single computer to work on assignments, having the system only registering one student. Similarly, interaction between these students cannot be measured this way (Pardo & Kloos, 2011). In our experiment, data from other resources was not aggregated with the data in the learning records store. For example, the virtual learning environment provides students the possibility to take quizzes. Quiz data (results, number of attempts, required amount of time to finish), however, are yet not stored and processed in the records store. In order to design and perform effective interventions, these data must be aggregated in future experiments.



### 3.4.4. Problem categorization based on learning analytics framework

In order to analyze which dimensions of the learning analytics framework (Greller & Drachsler, 2012) are causing problems during the implementation of the tool, we mapped the identified issues to the framework - see Table 10. This shows that five out of six dimensions faced difficulties so even though an almost 'plug-and-play' tool is provided, implementing it is a non-trivial endeavor.

| Dimension | Issues encountered |
|---|---|
| Stakeholders | Self-selection bias occurs as students need to opt-in for the experiment. IT-department not involved enough to collaborate fast on establishing external connection. |
| Objective | - |
| Data | Data on behavior in virtual learning environment alone do not provide enough insights to design effective pedagogical interventions. More data are needed in order to do so. |
| Instruments | The used tool was too limited in its data capturing. That is, not all activities were measured due to technical issues. |
| External limitations | Connection between case organization's and SURFnet's architecture only established at the very last moment. |
| Internal limitations | Teachers lacked insight in what learning resources to measure by placing tracking codes. |

**Table 10:** Problems Encountered per Dimension of Learning Analytics Framework

## 3.5. Future work

The experiment will be continued in the fall of 2017. At the Institute for Engineering and Design, two courses will implement the learning analytics tool in their course design. One of the courses is taught to fulltime students, the other one to part-time students. This provides the opportunity to explore behavioral differences between participants of the two programs. Furthermore, to date, no interventions were performed based on the data analysis and visualizations in the experiment. Learning analytics research in general often focuses on data collection, management or how data will help to improve education but designing effective pedagogical interventions becomes a critical element (Wise et al., 2016). Now we have demonstrated the tool can be implemented in the case organization's virtual learning environment, we aim to use it for intervention design. This calls for extended requirements, as described in Table 11. The next phase of the experiment will continue the current work and focus on answering the question in what way the visualizations in the learning analytics teacher dashboard can be used to design and perform timely pedagogical interventions. That is, interventions should take place when it is still possible to make changes to the learning behavior, i.e., during the course. Both the practical experiences gained and the to-be performed interventions benefit practitioners from the educational domain as it provides first-hand insights in the dynamics involved with starting learning analytics activities.

| Dimension | Value |
| --- | --- |
| Stakeholders | Students participating in one of two courses. Six lecturers teaching the courses, two of them actively involved in the experiment. SURFnet as external facilitator of the experiment. |
| Objective | Reflecting on learning activities of students and performing pedagogical interventions based on (the lack of) these activities. |
| Data | Activities of students in virtual learning environment, measured via the experimental tool. Formative assessment data. |
| Instruments | SURFnet's Learning Analytics Architecture, including teacher and student dashboards. (External) tools to capture additional data, e.g., formative assessment data. |
| External limitations | Privacy laws call students to opt-in for the experiment. Connection between case organization's and SURFnet's architecture. |
| Internal limitations | Time necessary to place tracking code on target pages and learning materials in the virtual learning environment. The teachers doing so must understand what data are required for the interventions. Competencies related to effective intervention design are required, that is, teachers must understand the visualizations and are able to design and perform useful interventions based on it. |

**Table 11:** Future Learning Analytics Experiment Case Description

3

# CHAPTER

# 4

# 4. From Dirty Data to Multiple Versions of Truth: How Different Choices in Data Cleaning Lead to Different Learning Analytics Outcomes[5]

*Learning analytics is the analysis of student data with the purpose of improving learning. However, the process of data cleaning remains underexposed within learning analytics literature. In this paper, we elaborate on choices made in the cleaning process of student data and their consequences. We illustrate this with a case where data was gathered during six courses taught via Moodle. In this data set, only 21% of the logged activities were linked to a specific course. We illustrate possible choices in dealing with missing data by applying the cleaning process twelve times with different choices on copies of the raw data. Consequently, the analysis of the data shows varying outcomes. As the purpose of learning analytics is to intervene based on analysis and visualizations, it is of utmost importance to be aware of choices made during data cleaning. This paper's main goal is to make stakeholders of (learning) analytics activities aware of the fact that choices made during data cleaning have consequences on the outcomes. We believe that there should be transparency to the users of these outcomes and give them a detailed report of the decisions made.*

## 4.1. Introduction

Virtual Learning Environments (VLEs) are digital learning platforms where students can interact with course materials (presentations, digital readers, instructional video's et cetera), can test their knowledge via quizzes, and can interact with each other and instructors via e.g., the discussion board. They support learning and simultaneously enable the collection of data on learner behavior in the system. Data from virtual learning environments are used for learning analytics activities, cf. Agudo-Peregrina, Iglesias-Pradas, Conde-González, & Hernández-García (2014); Conijn, Snijders, Kleingeld, & Matzat (2016); Rienties, Toetenel, & Bryan (2015); Romero, Ventura, & García (2008). Objectives of learning analytics vary but often involve student behavior modelling, prediction of performance and increase in (self) reflection and (self) awareness (Papamitsiou & Economides, 2014).

---

5 This work was originally published as: Knobbout, J.H., Everaert, H., and van der Stappen, E.J. (2019). From dirty data to multiple versions of truth: How different choices in data cleaning lead to different learning analytics outcomes. *BLED 2019 Proceedings*. Paper 57.

Importantly, raw data exported from virtual learning environments need to be cleaned and transformed before it is of any use to educators and students. In general, data cleaning takes up to 80% of analytical time (Brink, Richards, & Fetherolf, 2016). However, in the current learning analytics field, details about cleaning and transforming are often overlooked or, at best, not described and discussed in literature. For example, searching the terms *data cleaning* or *data preprocessing* in the Learning Analytics & Knowledge conference proceedings 2011 till 2018 (n = 438) only yield 17 papers describing either cleaning or preprocessing of learner data before analyzing the data. To make matters even more complex, full-scale and multimodal learning analytics require aggregated data from multiple sources, amplifying the effects of data cleaning on the analysis' outcomes. As we will show in this paper, data cleaning is problematic as (unspoken) choices can lead to a wide variety of outcomes and, subsequently, pedagogical interventions. Using a raw data set with VLE data, we will construct twelve different, cleaned sets and use these to calculate the time-spent-on the online part of six courses. With these data sets, we can provide an answer to our research question: "*What are the effects of (unspoken) choices made during the cleaning process of student data on the outcomes when these data are in turn used for learning analytics?*"



The remainder of this paper is structured as follows. First, an in-depth description of learning analytics and data cleaning is given based on existing literature. Then, the research question and method are described, followed by the presentation of our results. Finally, we provide five recommendations based on the outcomes of our study, as well as directions for future work.

## 4.2. Related work

In this section, we will present existing literature related to our study. First, we will provide a definition of learning analytics and an overview of the learning analytics process. Next, a thorough description of data cleaning and its implications is given.

### 4.2.1. Learning analytics

Learning analytics is "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environment in which it occurs" (Siemens et al., 2011). Learning analytics aim to improve learning processes at the level of students and teachers (Siemens & Long, 2011) and is, for example, used to analyze student behavior within digital learning environments, monitor the usage of course material, and predict whether students will fail a certain course or drop out entirely. The process of learning analytics consists of four steps: 1) learners generate learning data, 2) these data are captured, collected and stored, 3) analysis and visualization are performed, and 4) the design and use of data-driven pedagogical interventions (Clow, 2012) – see also Figure 12. Consequently, when the data is incorrect or incomplete, the analysis and subsequent interventions may be sub-optimal or even completely erroneous.

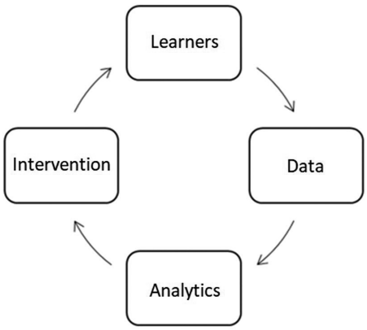


**Figure 12:** Learning Analytics Cycle (Clow, 2012).

### 4.2.2. Time-On-Task

Study-time is the time students spend on studying learning materials, using (metacognitive) tools, solving questions etcerera and can be used as measure of affected learning (Knobbout & van der Stappen, 2018). In several studies, a positive correlation between study-time and achievements of students has been found, cf. (Marzano, 2003; Scheerens & Bosker, 1997)). Estimating 'time-on-task' in the 'traditional' classroom is based on estimates by students and/or observations in classrooms. In a virtual learning environment (VLE), on the other hand, it is common to use the number of clicks (Wolff, Zdrahal, Nikolov, & Pantucek, 2013) or the time between certain clicks as measure for time-on-task (Kovanović et al., 2015).

Wolff et al. (2013) showed that "even fairly coarse grain data about students' activities" is useful in predicting retention (p. 148). Unfortunately, it is not perfectly clear what part of the clicks were used "[w]hile the issue of data cleaning for all data within the [Open University] was not resolved, it was possible to gain enough knowledge about the data [...] to start building models" (p. 146). From their point of view, it is import to note that in predicting failing students, changes in the student's own VLE activity, compared to their previous activity, are indicative. A relative reduction of clicks hints a failing student. Kovanović et al. (2015) deal explicitly and extensively with the thorny methodological issues of estimating time-on-task in VLE's. Their primary goal is "to raise awareness of the issue of accuracy and appropriateness surrounding time-estimation within the broader learning analytics community, and to initiate a debate about the challenges of this process" (p. 184). It is regarded good practice in different academic fields to discuss methodological issues and learning analytics should not become an exception to this rule. In this study, we extent the work of Kovanović et al. by estimating time-on-task for multiple parallel courses and by showing different options to handle missing data, i.e., records of events unlinked to any of the courses in the dataset.

### 4.2.3. Data cleaning

Data cleaning is an important part of the ETL (Extraction, Transformation and Load) process. According to VanderPlas (2016) the majority of the work in data science often "comprises cleaning and munging real-world data" (p. 188). Brink, Richards, and Fetherolf (2016) underline five common tasks, of which two - transforming original data to the target and create features that are more easily interpreted — are core business in working with large computer generated data files. Müller and Guido (2016) state that "in the real world, inconsistencies in the data and unexpected measurements are very common" (p. 19). Brink, Richards and Fetherolf (2016) estimate researchers are spending about 80% of their research time to munging, wrangling, combining or reshaping data. Special attention is given to utilizing expert knowledge. Although machine learning can reduce the need to create a set of expert-designed rules, that does not mean that prior knowledge of the application or domain should be discarded. Domain experts can help to identifying useful features that are more informative than the initial representation of the data (Müller & Guido, 2016).



### 4.2.4. Missing data

In (social sciences) papers and articles an often-subordinated subject is missing data. One of the most frequent and most ignored sources of bias is missing data (Baguley, 2012). Missing data is a stubborn problem in data analyses and, in general, we have to consider two issues: how much is missing and why it is missing. Thanks to eloquently written textbooks like 'Applied missing data analysis' (Enders, 2010), solutions to deal with missing data mechanisms are nowadays within reach for social researchers. In an overview of traditional techniques, Enders (2010) describes (listwise/pairwise) deletion, several imputation methods, averaging items in Likert scales, or last observation carried forward to address the problem and concludes that "most single case imputation methods produce biased estimates, even with Missing Completely at Random (MCAR) data. Stochastic regression imputation is the one exception and is the only traditional approach that yields unbiased estimates under a Missing At Random (MAR) mechanism" (p. 54). He demonstrates benefits of modern methods like maximum likelihood approaches and multiple imputation. Even in MCAR – which occurrence can hardly be safely assumed – the problems of missing data may become more serious if more cases are missing. "Unfortunately, there are as yet no firm guidelines for how much missing data can be tolerated for a sample of a given size" (Tabachnick & Fidell, 2007, p.63). Indirectly, Tabachnick and Fidell (2007) seem to consider about 5% missing or less of the sample size as 'manageable' in some way or the other. It also depends on the pattern of missing data. Choosing among different techniques for dealing with missing data may also depend on knowledge, confidence, and familiarity with the subject matter on part of the researcher. Van Belle (2011) among others advocates sensitivity analysis as a good idea based on "a thorough understanding of the subject matter" (p. 186).

It does not matter whether the above mentioned authors are working in the different fields varying form social or educational sciences, general data sciences to hard core machine learning and it seems fair to conclude that working with data is time consuming and in general comes with trouble, caveats or thorny issues. Fortunately, at the end of the process we will rely on some technical solutions, but working the data is in itself a muddy experience in which the data scientist/researcher has to rely on (several) subjective views and or decisions.

Educators are in the midst of a transition from learning analysis to learning analytics. The analysis of classical test scores is not enough. The availability of VLEs and the tracking of student behavior gives both students and educators much more opportunities to follow the learning of students in real-time and opportunities to intervene if necessary. At the same time, the upper limits of learning analytics are not well defined. Techniques borrowed from educational data mining, data science and machine learning combined with data from social-media become more and more intertwined (Daniel, 2017; Gibson & Ifenthaler, 2017). Technical solutions by themselves are not sufficient for successful use of educational data, as "[d]ata do not exist independently of the ideas, instruments, contexts and knowledge used to generate, process and analyze them" (Kitchin, 2014) (p. 2) thereby (implicitly) suggesting that data scientist are not aware of the pitfalls of data construction. As we will later show in this paper, most data scientists are aware of the true nature of data, that is, data are not neutral, objective and pre-analytic in nature. What often lacks is a thorough discussion of the possible solutions and consequences of a technical data issue, which is a major motive to conduct the study at hand.

## 4.3. Research method

The choices made in the cleaning of student data extracted from VLEs has effect on the outcome of this process – the dataset which is used for analysis and visualization of learning. However, not much is written about this effect and, consequently, the differences between outcomes based on the assumptions and choices made by the people responsible for the cleaning of the raw data are also underexposed. This study's aim is to fill this gap in the current learning analytics knowledge based on answering the following research question: *"What are the effects of (unspoken) choices made during the cleaning process of student data on the outcomes when these data are in turn used for learning analytics?"* As we will research how the made choices affect analytical outcomes of contemporary events whilst we do not have control over these events, a case study is a suitable research method for our study (Yin, 2013).



### 4.3.1. Case description

In this single case study, we analyze data from an international minor program. Students (n = 34) from the Netherlands, Finland, Spain, United Kingdom, Mexico, and Germany all participate in six blended courses (in this study named A to F), offered in 'traditional' classrooms, at an external workplace, as well as online via Moodle – a well-known VLE. In this study, we focus on data obtained from the latter.

Log files from Moodle are collected by exporting them via the administrator dashboard. This dashboard allows administrators to download all logs in comma separated value (.csv) format, which in turn can be processed in more specialized statistical software or learning analytics tools – in this study, we used IBM SPPS Statistics 24. The data are aggregated by us, i.e., events from all six courses are combined in one dataset. In compliance with the ethical procedures and guidelines that were applicable at the time the research was conducted, students were asked to give passive informed consent and all data were after collection immediately anonymized. Initially, the dataset comprises the variables as shown in *Table 12* and Figure 13.

| Variable | Description |
|---|---|
| Date | Date of the event taking place |
| Time | Time, in HH:MM-format, of the event taking place |
| User id | Moodle id of the user |
| Event context | Page of the VLE where event takes place |
| Component | Whether it involves an assignment or not |
| Event name | Name of the activity |
| Description | Description of the event, including course and user(s) id |
| Origin | Whether website or app is used |
| IP-address | IP-address from where Moodle is accessed |
| Id of affected user | In case of e.g., message sent or discussion board reaction |

**Table 12:** Variables extracted from Moodle.

| Date | Time | ID | Event context | Component | Event name | Description | Origir | IP addre |
|---|---|---|---|---|---|---|---|---|
| 6-9-2016 | 14:14 | 141 | Assignment: Task 2) | Assignment | The status of the su | The user with id '141' has viewed the submission status page for the as | web | 145.89.118. |
| 6-9-2016 | 14:14 | 141 | Course: Introductory | System | Course viewed | The user with id '141' viewed the course with id '16'. | web | 145.89.118. |
| 6-9-2016 | 14:11 | 118 | Course: Introductory | System | Course viewed | The user with id '118' viewed the course with id '16'. | web | 145.89.164. |
| 6-9-2016 | 14:11 | 33 | Course: The Project - | System | Course viewed | The user with id '33' viewed the course with id '17'. | web | 145.89.64.1 |
| 6-9-2016 | 14:08 | 118 | Assignment: Task 0t | Assignment | The status of the su | The user with id '118' has viewed the submission status page for the as | web | 145.89.164. |
| 6-9-2016 | 14:08 | 118 | Assignment: Task 0t | Assignment | A submission has l | The user with id '118' has submitted the submission with id '580' for the | web | 145.89.164. |
| 6-9-2016 | 14:08 | 44 | User: | System | Message viewed | The user with id '44' read a message from the user with id '118'. | web | 145.89.164. |
| 6-9-2016 | 14:08 | 118 | System | System | Message sent | The user with id '118' sent a message to the user with id '44'. | web | 145.89.164. |
| 6-9-2016 | 14:08 | 38 | User: | System | Message viewed | The user with id '38' read a message from the user with id '118'. | web | 145.89.164. |
| 6-9-2016 | 14:08 | 118 | System | System | Message sent | The user with id '118' sent a message to the user with id '38'. | web | 145.89.164. |
| 6-9-2016 | 14:08 | 46 | User: | System | Message viewed | The user with id '46' read a message from the user with id '118'. | web | 145.89.164. |

**Figure 13:** Snippet of raw data set.

As a case for our study, we want to determine for each individual student how much time is spent on each of the six courses of the minor program and the underlying learning activities. This means we have to structure the data in such way that we can estimate the time-on-task for all events in the data set. We elaborate on this process and its results in the next section.

### 4.3.2. Cleaning of the data

Our focus in the ETL process of the Moodle data is on cleaning and transforming the data by deriving new calculated variables and values by splitting a column (existing variable) into multiple columns (new variables) and so disaggregating the data. Our VLE data records user id, event description and timing of an event. The variable Description (including the user id and course id) is split in different variables to identify the course the student is working on. We are willing to assume that a student's action in the VLE and thus creating an event in the data set is synonymous with studying. Therefore, we have to assume that opening of a second event implies the end of the first event and the time-spent-on the first event T1 amounts to t2 minus t1 – see Figure 14. Unfortunately, closing of the event is normally not registered in the VLE. Consequently, time-spent-on the last event in a session (T4 in Figure 14) cannot reliably be calculated.

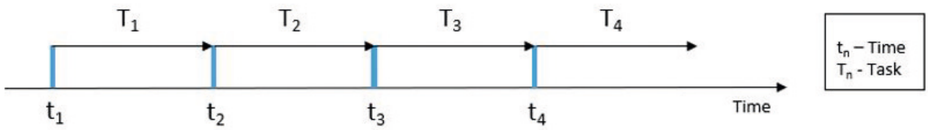


**Figure 14:** Calculation of time-spent-on task by using the start of new event.

Another issue is missing data: many events are not linked to a specific course. For example, when a student sends a message to another student, Moodle does not know to what course (if any at all) the message relates and therefore omits the inclusion of a course id in the event description. This proves problematic when calculating the total-time-spent-on a course. In Figure 15 we see that a student is working on course D at t2. Later, at t5, he is involved in course C. In order to link the other events (t1, t3, t4, t6, and t7) to a specific course to compute total-time-spent-on a course, we must make some assumptions.

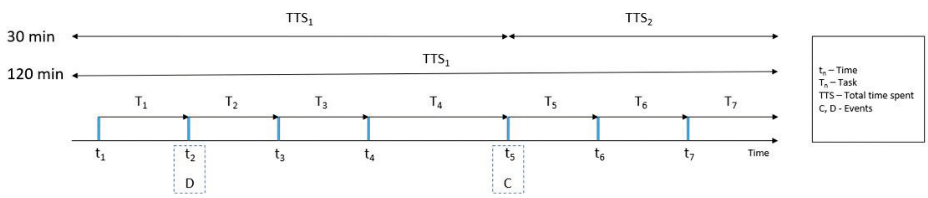


**Figure 15:** Total-time-spent on different courses, based on varying session times.

First, we must decide whether the event on t4 is to be associated with a session in which the student is working on course C or course D. In the literature, a session or study-period often ends 30 minutes after the last click (see discussion and overview of time-on-task in (Kovanović et al., 2015)). Moodle's default setting, however, automatically ends sessions after 120 minutes. That are two main versions we worked with in this study, but there is no logical reason to limit ourselves to these options – why not 60 or 90 minutes? By deciding to end a session after 30 minutes of inactivity, we also assumed that the course worked on in the 30 minutes version is D at t1, t2, t3 and t4, while the student started with course C at t5. We can now calculate the total-time-spent (TTS) during this session by adding all Tx within the session. In the default Moodle version, on the other hand, the timing between all events is smaller than the 120 minutes cut-off time. In such a study period (see Figure 15), we can calculate the total-time-spent during the session but do not know to what (portion of a) course to assign it. It can be DDDDCCC, but also DDCCCCC or whatever permutation possible. Obliviously, this is of influence when computing total-time-spent-on a course.

To deal with the problem of events not linked to courses – which is essentially a missing data issue – we defined six scenarios:

- In the first scenario (strict) we disregarded sessions with events not referring to any course. This way, we do not have to make assumptions to what course a session relates. The downside, however, is that we lose sessions and, thus, information.
- In the second scenario (wide 1), we filled out the missing values by carrying the last observation forward till the next observed course or the end of the study session.
- In the third scenario (wide 2), we simply relied on the most frequent course in a study period as the one and only; overwriting missing values in that particular time frame.

In the other three scenarios, we imputed the missing values with randomly assigned courses weighted by the number of known courses worked on:

- In the fourth scenario (wide 3), the weight was based on the number of all courses observed on a weekly basis of all students together and all missing values of a single student in a particular time frame got the same random course assigned (for instance, AAA or BBB)
- In the fifth scenario (wide 4), the same is done as in wide 3 but several missing values in a particular computed study-period were independently randomized (for instance, DBA, or CAC or just FFF).
- In the final scenario (wide 5), the weight is computed by the number of courses directly chosen by an individual student on a weekly basis and missing values were imputed as in wide 4.

We just want to show that all scenarios are plausible in one way or the other, and indeed, we could have chosen other ways to deal with missing values. At this point we are not interested in the stability of the different approaches. In order to compute the total-time-spent-on a course (TTSA, TTSB et cetera) in the different versions and session, we recomputed the study sessions by taking t-last minus t-first of a row of equal courses in order to estimate time-spent-on a course. See Figure 16 for a schematic representation of some of the scenarios.

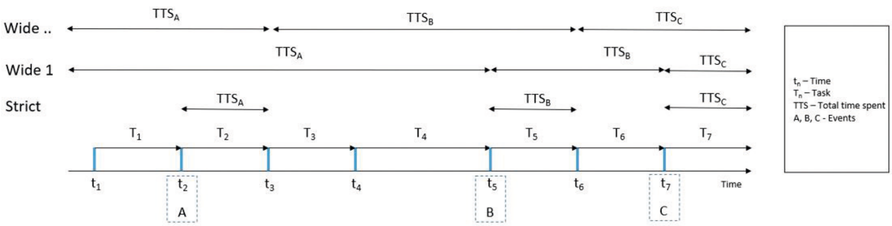

**Figure 16:** Schematic representation of scenarios Strict, Wide 1 and one of the other Wides.

### 4.3.3. Data processing

In line with our own recommendations (see section 4.5.1), we provide a summary of assumptions and decisions made in the processing of our data:

- Events related to accessing the VLE with phones or mobile apps creates records without any information other than that a mobile device is used and can be removed from the dataset;
- Activities as changing passwords or failed login attempts are not related to learning and thus can be removed from the dataset;
- Our research focuses on learners so event caused by other users (teachers, administrators etcetera) can be removed from the dataset;
- All remaining events in the dataset represent learning activities in the VLE;
- Learning sessions end either 30 or 120 minutes after the start of the last event in said session;
- Data is cleaned by applying one of the six methods described in section 4.3.2.

## 4.4. Results

Now we have 12 different data sets – the six scenarios how to deal with missing data and two different sessions times (30 versus 120 minutes). With these data sets, we now calculate the time-spent-on the six courses of the minor program.

### 4.4.4. Identifying events and courses

In total, our raw dataset comprised 148,285 events. After removing events related to accessing the VLE with phones or mobile apps removing non-learning activities, and limiting ourselves to student users, we end up with 57,811 events. Of all these events, just 12,334 events (21% of relevant events) are directly linked to a course – see Figure 17. This leaves 45,477 events (79%) unaccounted for and the only way to link the registered student activity to a course is within a study session based on the Moodle default of 120 minutes or the 30 minutes often used in academic studies.

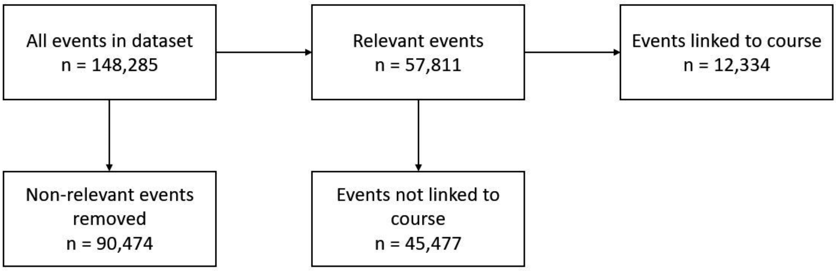


**Figure 17:** Number of events during and after data processing.

As a result of the option between 30 and 120 minutes, we see in Table 13 that in the 120 minute default 3,832 events take place within study periods in which there is no link to any course at all. Just by shortening the end of the study session to 30 minutes, the number of not directly identifiable events more than doubles to 8,546 events. Shorter periods in the 30 minutes version leads to more unequivocally identifiable events; all known events in these periods belong to one and the same course. In the 120-minute default, it is just the opposite: the number of events pertaining to two or more different courses within a study period nearly doubles compared to the 30 minutes variant. Independent of the selected version, there are 137 not directly identifiable events we could not solve by carrying the last observation forward till the next observed (wide 1) or just taking the most frequent course in a study session (wide 2).

### 4.4.5. Identifying time-spent-on tasks and courses

After cleaning the data and imputing the missing values, we have 12 datasets and can calculate the number of activities on each course based on the various data sets. At first glance it seems that only differences between the strict and the wide scenarios are noteworthy. The solutions within the five wide approaches do not differ that much. That is erroneous: the number of events in Table 13 are presented over all students together. What we really want to know is the number of events – and more importantly – time-spent-on by each individual student. Both measures vary enormously according to the chosen dataset. We can now also calculate the time-spent-on each course by each individual student as shown in Table 14 for just four students.

| Moodle default: 120 minutes | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Decision** | | **Unidentified** | A | B | C | D | E | F | **Total events** |
| Raw data | | | | | | | | | 148,285 |
| Evens relating to courses | Unidentified | **3,832** | 0 | 0 | 0 | 0 | 0 | 0 | 3,832 |
| | Unequivocally identified | 0 | 15,125 | 6,005 | 3,651 | 1,126 | 1,943 | 5,744 | 33,594 |
| | Two or more options | | 7,407 | 8,182 | 6,563 | 8,127 | 9,624 | 9,425 | |
| Versions | Strict | 3,832 | 18,315 | 10,112 | 5,944 | 3,971 | 5,115 | 10,522 | 53,979 |
| | Wide 1 | 137 | 19,058 | 10,552 | 6,329 | 4,514 | 5,701 | 11,520 | 57,674 |
| | Wide 2 | 137 | 19,117 | 10,788 | 6,175 | 4,126 | 5,387 | 12,081 | 57,674 |
| | Wide 3 | 0 | 18,902 | 10,768 | 6,242 | 4,303 | 5,620 | 11,976 | 57,811 |
| | Wide 4 | 0 | 18,885 | 10,784 | 6,239 | 4,314 | 5,621 | 11,968 | 57,811 |
| | Wide 5 | 0 | 19,017 | 10,653 | 6,277 | 4,409 | 5,635 | 11,820 | 57,811 |
| **Theoretical standard/advise: 30 minutes** | | | | | | | | | |
| **Decision** | | **Unidentified** | A | B | C | D | E | F | **Total events** |
| Raw data | | | | | | | | | 148,285 |
| Evens relating to courses | Unidentified | **8,546** | 0 | 0 | 0 | 0 | 0 | 0 | 8,546 |
| | Unequivocally identified | 0 | 15,560 | 7,080 | 4,005 | 2,009 | 2,970 | 7,285 | 38,909 |
| | Two or more options | | 4,641 | 4,371 | 3,862 | 4,075 | 3,984 | 3,529 | |
| Versions | Strict | 8,546 | 17,584 | 9,663 | 5,276 | 3,430 | 4,624 | 8,688 | 49,265 |
| | Wide 1 | 137 | 19,051 | 10,562 | 6,351 | 4,513 | 5,718 | 11,479 | 57,674 |
| | Wide 2 | 137 | 19,116 | 10,827 | 6,143 | 4,357 | 5,758 | 11,473 | 57,674 |
| | Wide 3 | 0 | 18,784 | 11,068 | 6,207 | 4,227 | 5,616 | 11,909 | 57,811 |
| | Wide 4 | 0 | 18,827 | 11,138 | 6,145 | 4,148 | 5,782 | 11,771 | 57,811 |
| | Wide 5 | 0 | 18,989 | 10,963 | 6,221 | 4,316 | 5,669 | 11,653 | 57,811 |

**Table 13:** Assigning events to courses in order to estimate time-spent-on course.

| | | Moodle default: 120 minutes | | | | | | | Theoretical standard/advise: 30 minutes | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| User id | Scenario | A | B | C | D | E | F | Total time (minutes) | A | B | C | D | E | F | Total time (minutes) |
| 128 | Strict | 16% | 26% | 18% | 13% | 10% | 17% | 14.173 | 23% | 28% | 21% | 11% | 5% | 11% | 3.439 |
| | Wide1 | 15% | 24% | 16% | 12% | 9% | 24% | 15.921 | 21% | 26% | 20% | 10% | 5% | 19% | 3.999 |
| | Wide2 | 13% | 26% | 13% | 12% | 6% | 30% | 17.261 | 19% | 28% | 17% | 12% | 4% | 19% | 4.116 |
| | Wide3 | 15% | 27% | 16% | 12% | 11% | 20% | 15.921 | 20% | 28% | 19% | 12% | 6% | 15% | 3.999 |
| | Wide4 | 15% | 25% | 16% | 12% | 9% | 23% | 15.397 | 22% | 27% | 21% | 11% | 5% | 15% | 3.747 |
| | Wide5 | 14% | 24% | 15% | 13% | 11% | 23% | 16.224 | 21% | 28% | 21% | 10% | 5% | 15% | 3.982 |
| 132 | Strict | 28% | 13% | 18% | 7% | 9% | 26% | 6.466 | 44% | 9% | 10% | 4% | 12% | 20% | 1.447 |
| | Wide1 | 22% | 18% | 13% | 8% | 7% | 32% | 9.142 | 32% | 7% | 8% | 12% | 10% | 32% | 2.070 |
| | Wide2 | 21% | 17% | 12% | 10% | 9% | 30% | 9.618 | 32% | 7% | 8% | 11% | 11% | 32% | 2.105 |
| | Wide3 | 22% | 13% | 13% | 8% | 11% | 32% | 9.142 | 32% | 14% | 8% | 8% | 9% | 29% | 2.070 |
| | Wide4 | 25% | 13% | 15% | 7% | 9% | 32% | 8.017 | 37% | 8% | 9% | 5% | 11% | 30% | 1.795 |
| | Wide5 | 19% | 10% | 12% | 7% | 20% | 33% | 10.387 | 33% | 7% | 8% | 5% | 21% | 27% | 2.106 |
| 138 | Strict | 26% | 21% | 13% | 3% | 11% | 27% | 5.385 | 45% | 11% | 5% | 3% | 7% | 29% | 1.424 |
| | Wide1 | 20% | 22% | 13% | 4% | 11% | 31% | 7.008 | 32% | 11% | 13% | 2% | 8% | 34% | 2.011 |
| | Wide2 | 21% | 20% | 13% | 6% | 6% | 34% | 7.157 | 35% | 11% | 10% | 3% | 8% | 34% | 2.035 |
| | Wide3 | 20% | 22% | 11% | 2% | 13% | 32% | 7.008 | 34% | 11% | 11% | 5% | 7% | 32% | 2.011 |
| | Wide4 | 22% | 25% | 12% | 2% | 9% | 29% | 6.226 | 36% | 13% | 10% | 2% | 5% | 33% | 1.804 |
| | Wide5 | 21% | 27% | 14% | 3% | 10% | 25% | 6.658 | 28% | 23% | 10% | 3% | 5% | 31% | 2.325 |
| 144 | Strict | 44% | 6% | 7% | 9% | 13% | 22% | 4.577 | 45% | 10% | 12% | 9% | 1% | 23% | 1.336 |
| | Wide1 | 29% | 8% | 9% | 23% | 8% | 23% | 7.759 | 36% | 8% | 11% | 24% | 0% | 21% | 1.802 |
| | Wide2 | 26% | 6% | 9% | 8% | 26% | 25% | 8.166 | 34% | 10% | 10% | 24% | 1% | 21% | 1.829 |
| | Wide3 | 27% | 6% | 6% | 10% | 10% | 41% | 7.759 | 36% | 9% | 9% | 9% | 5% | 32% | 1.802 |
| | Wide4 | 31% | 5% | 7% | 7% | 16% | 34% | 6.674 | 41% | 9% | 11% | 7% | 1% | 31% | 1.580 |
| | Wide5 | 37% | 8% | 4% | 7% | 15% | 29% | 6.787 | 42% | 10% | 12% | 8% | 1% | 28% | 1.526 |

**Table 14:** Relative amount of time spent on courses for four different students.

4

Compared to the theoretical standard/advice of 30 minutes, students spend about 3 to 4 times as much time on the total of six courses under the Moodle default of 120 minutes. Considering Moodle's default session ending time of 120 minutes, students spent about 3 to 4 times as much time on their courses compared to the total time-on-task when using the theoretical standard ending time of 30 minutes. This is in line with the assumptions used – 120 minutes is four times as long as 30 minutes. However, if we look at the relative time students spent on specific courses between the two versions or within the used scenarios of a version, the link between assumptions used and relative time becomes foggy and blurred.

In the 30 minutes version, all students seem to spend relatively more time on course A and less on course B, compared to the Moodle default of 120 minutes. However, student 144 spends also relatively less time on course E. If we compare over the scenarios within the separate versions, we sometimes see huge differences between strict and several wide scenarios. For instance, in the 30 minutes version, student 138 spends 45% of his time in the strict version to course A, in wide 5 this is reduced to a mere 28%. In the Moodle default, the relative time-spent-on in these sets is more or less the same (26% versus 21%).

As our results show, it is difficult to see a common pattern in these figures, indicating different assumptions lead to different dashboard figures. Concluding, we observe that time-spent-on as a key variable for the quality of learning stays without reach for teachers as a basis to act upon and interfere with a particular student: it just depends and variates with the assumptions made and the truth is hard to find.

## 4.5. Discussion and conclusion

In this paper, we have shown that the choices made during the cleaning process of student data can have large impact on the outcome of the subsequent analysis. Estimating time-on-task is one example of a learning (outcome) measure which is affected by data cleaning, but also other metrics used in learning analytics research might be influenced, e.g., the use of (metacognitive) tools or the number of discussion board postings. With the emerge of full-scale and multimodal learning analytics – requiring the aggregation of data from multiple sources –the effects of data cleaning on the analysis' outcomes are even more amplified. We are not in search of a holy grail for student data cleaning (which probably does not exist at all), but the goal of this study is to make both practitioners and academics aware of these - often unspoken - choices and their effect.

### 4.5.1. Recommendations

Based on our research, we present the following recommendations: (1) provide users of learning analytics tools (students, teachers et cetera) with the insight what assumptions and corresponding choices were made during the data cleaning process. This helps them to better understand the results and visualizations of the data analysis; (2) provide users with the opportunity to see other versions based on different assumptions of the data set as well; (3) to make scientific work better reproducible and comparable, researchers should elaborate on the cleaning of their data. In the current literature, researchers often almost immediately jump from raw data to results without saying anything on the choices made, although some exceptions exist, cf. Bos and Brand-Gruwel (2016); Chen, Chen, and Xing (2015); Kovanović et al. (2016); (4) involve domain experts in the cleaning process. Data experts working on the data sets without knowing the exact context the data was collected in, might use erroneous assumptions to clean the data. By consulting domain experts before the data handling, the resulting data might be better suit the learning context (Müller & Guido, 2016); (5) stakeholders should feel responsible, support the choices made, and be transparent about them.

If we want students, colleagues and other professionals to work with our analysis, results or dashboard functionality, we should be open and give them a detailed report of the decisions made. As a rule of thumb, we should state and explain explicitly how we have dealt with the issues at hand in such way the user can understand it (Van Belle, 2011).

### 4.5.2. Future work

Now we have different data sets, we might want to research in what ways to inform end users about the data cleaning process. That is, how can we inform users – students, teachers et cetera – what assumptions were made, what steps were taken, what user preferences are, and what the effects on the analysis outcome are. We propose the use of focus groups to identify (critical) success factors for awareness creation about data cleaning and its consequences.

# PART III

## Design

# CHAPTER

# 5

# 5 A Capability Model for Learning Analytics Adoption: Identifying organizational capabilities from literature on learning analytics, big data analytics, and business analytics[6]

*Despite the promises of learning analytics and the existence of several learning analytics implementation frameworks, the large-scale adoption of learning analytics within higher educational institutions remains low. Extant frameworks either focus on a specific element of learning analytics implementation, for example, policy or privacy, or lack operationalization of the organizational capabilities necessary for successful deployment. Therefore, this literature review addresses the research question "What capabilities for the successful adoption of learning analytics can be identified in existing literature on big data analytics, business analytics, and learning analytics?" Our research is grounded in resource-based view theory and we extend the scope beyond the field of learning analytics and include capability frameworks for the more mature research fields of big data analytics and business analytics. This paper's contribution is twofold: 1) it provides a literature review on known capabilities for big data analytics, business analytics, and learning analytics and 2) it introduces a capability model to support the implementation and uptake of learning analytics. During our study, we identified and analyzed 15 key studies. By synthesizing the results, we found 34 organizational capabilities important to the adoption of analytical activities within an institution and provide 461 ways to operationalize these capabilities. Five categories of capabilities can be distinguished – Data, Management, People, Technology, and Privacy & Ethics. Capabilities presently absent from existing learning analytics frameworks concern sourcing and integration, market, knowledge, training, automation, and connectivity. Based on the results of the review, we present the Learning Analytics Capability Model: a model that provides senior management and policymakers with concrete operationalizations to build the necessary capabilities for successful learning analytics adoption.*

6 This work was originally published as: Knobbout, J.H., and van der Stappen, E.J. (2020). A Capability Model for Learning Analytics: Identifying Organizational Capabilities from Literature on Learning Analytics, Big Data Analytics, and Business Analytics, in *International Journal of Learning Analytics and Artificial Intelligence for Education*, 2. 1.

## 5.1. Introduction

Learning analytics aim at optimizing learning and the environment in which learning occurs by analyzing and intervening on learner-generated data (LAK, 2011). Although the results show promising effects, much learning analytics practice in the past decade is done at a small scale with a limited number of students and teachers involved. As a result, examples of the large-scaled application within higher educational institutions remain scarce. Learning analytics can bring competitive advantages to the educational domain, but to do so, institutions must invest in resources and institutional capacities (Arnold, Lynch, et al., 2014). This investment requires strategic planning at the highest level of the institution. To address the strategic investment higher educational institutions need to make, we take the lens of the resource-based view theory as our main perspective. The resource-based view has been used to study, among others, big data analytics and business analytics – two research fields similar to learning analytics. Hence, we consider it useful to the learning analytics community and use this theory to study learning analytics adoption.



This study aims to identify organizational capabilities for large-scale implementation and adoption of learning analytics in higher educational institutions. As we want to aggregate the findings of prior studies to develop a new model, we conduct a literature review. Therefore, our paper has two main contributions to the research field. The first is a literature review on the commonalities and differences between capabilities for business analytics, big data analytics, and learning analytics. The second is a capability model to support the implementation and uptake of learning analytics. We enhance the current body of knowledge by not only providing an overview of important capabilities but also their operationalization. This important aspect is often overlooked in existing models on learning analytics implementation. Moreover, rather than limiting ourselves to the field of learning analytics, in our search we include literature from research fields with a longer history of using data to enhance processes and the environment in which these processes take place. In contrast to existing models, we take a comprehensive look at the implementation and adoption rather than only a specific part of it like privacy and ethics (e.g. (Greller & Drachsler, 2012; Hoel et al., 2017; Pardo & Siemens, 2014)) or policy (e.g., Ferguson et al., 2014a; Tsai et al., 2018a). Finally, to the best of our knowledge, we are the first ones who use the resource-based view to study learning analytics. The review provides an answer to the main research question: "*What capabilities for* *the successful adoption of learning analytics can be identified in existing literature on big data, business and learning analytics?"*

The remainder of this paper is structured as follows. First, we provide an overview of the background of the study. We will then describe in detail the methodology we applied, after which we present the results of our study. Finally, in the discussion section, we provide recommendations for future work, including the planned approach for refinement and validation of the Learning Analytics Capability Model and as well as a discussion on the limitations of our study. The complete set of ways to operationalize the learning analytics capabilities is published online[7].

## 5.2. Theoretical background

In this section, we start with an overview of known problems faced by higher educational institutions when trying to adopt learning analytics at scale. We then describe some of the known frameworks supporting the uptake of learning analytics by higher educational institutions. Finally, we elaborate on the resource-based view.

### 5.2.1. Learning analytics adoption challenges and frameworks

Much research focuses on the application of learning analytics in a limited context (Gašević et al., 2016). As a result, the institutional adoption of learning analytics and embedding in educational systems remains quite immature (Colvin et al., 2015; Gašević et al., 2019; Wise & Vytasek, 2017). A systematic literature review by Viberg et al. (2018) on the use of learning analytics in higher education shows that 94% of the studies described in the reviewed papers (n=252) does not scale. A reason for this might be that higher educational institutions scaling up on learning analytics face a variety of problems and challenges, e.g., issues with usability, access, performance, and calculation (Lonn et al., 2013), concerns about privacy and ethics (Drachsler & Greller, 2016), lack of exemplars and guiding resources as well as technical, social, and cultural issues (Dawson et al., 2018), or proving the value of learning analytics, aligning it with learning sciences, and collecting useful data in a secure way (Nouri et al., 2019). In a review of extant literature, Tsai and Gašević (2017b) identified six primary challenges related to strategic planning and learning analytics policies, including a shortage of leadership capabilities and insufficient training opportunities for end-users. Empirical research by Ifentaler

7 https://www.researchgate.net/publication/339847879_Learning_Analytics_Capability_Model

& Yau (2019) shows that stakeholders often can identify the resources necessary for learning analytics adoption but that most institutions still need to build and attain these required resources.

The issues and challenges withholding higher educational institutes to adopt learning analytics successfully attract the attention of scholars. Noticeable studies on the subject of learning analytics implementation are the Europe-oriented Supporting Higher Education to Integrate Learning Analytics (SHEILA) framework (Tsai et al., 2018) and its Latin American counterpart, the LALA framework (Maldonado-Mahauad et al., 2018). Both frameworks can be used to inform strategic planning and policy processes for large-scale implementation in higher education contexts. The SHEILA framework's focus is on policy development and comprises six dimensions, each containing three key elements. Although questions in the framework prompt answers and actions which help institutions to mitigate challenges, policies do not necessarily provide direct solutions to the identified challenges (Tsai et al., 2018). The LALA framework, which is highly influenced by the SHEILA framework, is composed of four fundamental dimensions (Pérez-Sanagustín et al., 2019). The framework is yet grounded in theory and empirical validation is suggested as future work. Nonetheless, preliminary results show that there is no such thing as "one-size-fits-all" for large-scale learning analytics adoption, as institutional needs differ per university. During a literature review, Colvin et al. (2017) identified nine different frameworks to support learning analytics implementation. From their analysis, it can be learned that five dimensions are considered to impact implementations: technological readiness, leadership, organizational culture, staff and institutional capacity, and learning analytics strategy. However, the authors state that "operationalizations of these dimensions varied across the literature" (Colvin et al., 2017, p. 285).

To the best of our knowledge, there is no literature review conducted to identify and analyze the different ways organizational capabilities supporting the adoption of learning analytics are operationalized. Our study aims to fill this knowledge gap. Successful adoption is not only about possessing the right resources (e.g., hardware, software, skilled people) but also about the ways these resources are deployed and managed. This is best described by the resource-based view, which we introduce in the next paragraph.

### 5.2.2. Resource-Based View

The resource-based view attributes organizational performance to its resources, which, to obtain sustained competitive advantages, must be valuable, rare, inimitable, and non-substitutable (Barney, 1991; Bharadwaj, 2000). They are generally divided into categories as financial resources, physical resources, human resources, technological resources, organizational resources, and reputation (Barney, 1991; Grant, 1991). Resources relate to assets and capabilities (Helfat & Peteraf, 2003; Wade & Hulland, 2004). Assets involve anything which can be deployed by organizations to create, produce and offer its goods or services to a market and can be either be tangible, intangible or personnel-based (Bharadwaj, 2000). Capabilities, on the other hand, are repeatable patterns of actions in the use of these assets (Wade & Hulland, 2004). They involve “complex patterns of coordination between people and between people and other resources” (Grant, 1991, p. 122) and are essentially interacting routines. Capabilities are a special kind of resource since they refer to an organization’s capacity to deploy other resources and ownership cannot be transferred between organizations. Capabilities are strongly “embedded in the organization and its processes” (Makadok, 2001, p. 388) and cannot easily be bought, but need to be built in order to effectively interact with the organizational processes and procedures.

To research what capabilities are necessary for learning analytics, we turn to two adjacent research fields: big data analytics and business analytics. With ever-growing datasets - both in size and complexity - big data analytics provides the required knowledge about “advanced and unique data storage, management, analysis, and visualization technologies” to handle these datasets (H. Chen et al., 2012, p. 1166). Business analytics, on the other hand, analyze data to understand and manage businesses more effectively (Kohavi et al., 2002) and is parallel to analytics in an educational setting (Barneveld et al., 2012). The resource-based view has been used to study capabilities for big data analytics and business analytics in the past (Adrian et al., 2018).

As the field of learning analytics is younger than big data analytics and business analytics, we choose to apply exaptation. Exaptation is the process of extending known solutions in one domain to solve problems in another domain (Gregor & Hevner, 2013). These solutions have a high degree of maturity in one domain but the application maturity in the focal domain is yet low. Consequently, prior ideas need to be tested and refined, resulting in opportunities for research and knowledge contribution. Although learning analytics have a specific goal – improve learning and learning environments – the goals and intents of analytics at the institutional level are similar for organizations in the educational domain and those in other domains. Therefore, in our research, we not only look at learning analytics literature but include studies from big data analytics and business analytics as well.

## 5.3. Methodology

In this study, we will answer the main research question: "*What capabilities for the successful adoption of learning analytics can be identified in existing literature on big data analytics, business analytics, and learning analytics?*" The following sub-questions operationalize the main research question:

RQ1: "*What capabilities necessary for the successful adoption of big data analytics and business analytics within an organization can be identified in existing literature?*"

RQ2: "*What capabilities necessary for the successful adoption of learning analytics within a higher educational institution can be identified in existing literature?*"

RQ3: "*Which similarities and differences can be identified between capabilities for big data analytics and business analytics, and learning analytics?*"

5

Figure 18 shows the relationship between the main research question, the sub-questions and the final outcome of the study: the Learning Analytics Capability Model.

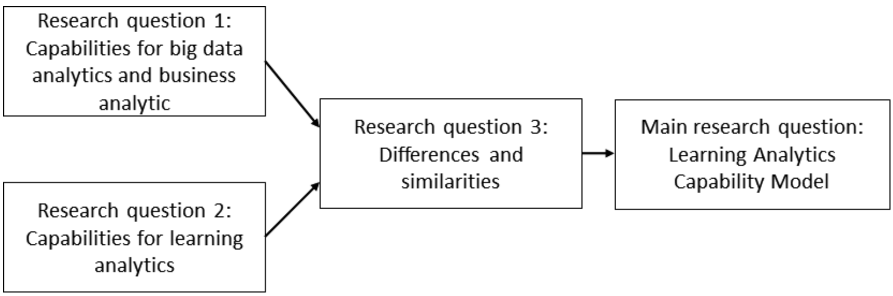


**Figure 18:** Relationship between the main research question and sub-questions.

### 5.3.1. Method for RQ1

Much research towards the required capabilities for big data analytics and business analytics has already been performed. Adrian et al. (Adrian et al., 2018) have conducted a systematic literature review to investigate factors and elements affecting big data analytics implementation while taking a resource-based perspective. The authors identified 15 key studies, which we will initially include in our research. As we also want to include literature on business analytics capabilities and are interested in the way capabilities can be operationalized, we conduct an additional literature review. We are particularly looking for papers developing capability frameworks, for these extensively describe both capabilities and their operationalization. As we want to include literature from many different domains, we use Google Scholar as search engine. We use the following search string: *("big data analytics capabilities" OR "big data analytics capability" OR "business analytics capabilities" OR "business analytics capability" OR "BA capability" OR "BDA capability" OR "BA capabilities" OR "BDA capabilities") AND ("resource-based view")*. To select key studies for analysis, we apply the inclusion and exclusion criteria shown in Table 15.

| **Criterion** | **Inclusion** | **Exclusion** |
|---|---|---|
| Language | English | Non-English |
| Outlet | Peer-reviewed conference proceeding papers or journal papers | Book (chapters), master thesis, editorial comments |
| Framework | Research on big data analytics and business analytics capability frameworks | Research on individual capabilities or anecdotal research findings |
| Operationalization | Provides a description of the operationalization of capabilities | No operationalization provided |
| Validation | Empirically validated frameworks | No validation |
| Citations | Cited by others at least once | Not cited by others |
| No follow-up | Newly identified framework | Follow-up studies using already identified framework |

**Table 15.** Inclusion and exclusion criteria.

Based on titles and abstracts, papers not meeting our selection criteria are removed from the dataset. Next, by reading the full texts of the remaining papers, key studies are identified. From the key studies, the operationalizations of analytical capabilities are extracted and coded based on open coding principles. In open coding, items are compared with each other for similarities and then labeled, allowing conceptually similar items to be grouped to form categories (Strauss & Corbin, 1990). Capabilities can variate in level, resulting in a hierarchical order (Ambrosini et al., 2009). In our study, we distinguish between third-order, second-order, first-order, and zero-order capabilities. Third-order capabilities are the highest level and describe the core concept. Second-order capabilities describe the different categories of capabilities within the core concept. First-order capabilities describe the abilities necessary to achieve individual tasks. Finally, zero-order capabilities are the ways first-order capabilities are operationalized – see Figure 19. This leveling will be used to structure the outcomes of our literature review. Based on similarity, we group operationalizations into first-order capabilities, which in turn are categorized into second-order capabilities. To secure the quality of the coding process, all coding is done by two researchers in parallel. The results are compared and any differences are discussed until consensus is reached.

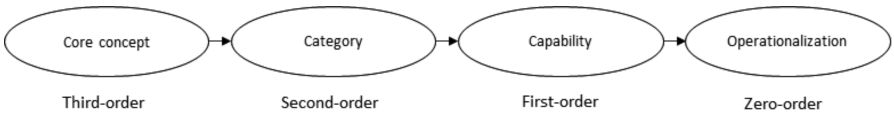


**Figure 19:** Hierarchical order of various levels of capabilities.



### 5.3.2. Method for RQ2

In a recent review of existing literature on learning analytics deployment, Colvin et al. (2017a) identified a dozen learning analytics implementation models. We take this study as the starting point for our second research question and include the 12 studies in our search process. To make sure no relevant models are missed, we perform an additional search in two major databases in which, among others, papers from the Journal of Learning Analytics and the Learning Analytics and Knowledge (LAK) conference proceedings papers are published: Education Resources Information Center (ERIC) and Association for Computing Machinery (ACM). We use the search string *“learning analytics” AND (adoption OR uptake OR implementation) AND (capability OR capacity OR process OR routine OR asset OR “resource-based view”)* for both databases. On the models identified by Colvin et al. and the papers we found during the additional search, the same criteria as for research question 1 are applied (Table 15) with only one exception. Instead of describing research on big data analytics and business analytics capability frameworks, papers must describe research on learning analytics implementation, adoption, and/or use at scale. Titles and abstracts are scanned to remove papers clearly not meeting the inclusion criteria. The final selection of key studies will be made by thoroughly reading the full texts of the remaining papers and comparing them with the selection criteria.

In the first round of coding, the operationalizations extracted from the key studies are coded based on the a priori coding scheme: the outcomes of research question 1. That is, the capabilities defined in that part of our study are used to identify similar capabilities in the learning analytics frameworks. Concepts not relating to any of these capabilities are then coded based on open coding principles (Strauss & Corbin, 1990). This way, we can identify capabilities and operationalizations unique for learning analytics compared to big data analytics and business analytics. Similar to the coding process for the first research question, the coding will be done by two researchers who code, compare and discuss all capabilities found during the search.

### 5.3.3. Method for RQ3

The first two research questions lead to data on the capabilities for either big data analytics and business analytics or learning analytics. In the third research question, differences and similarities between the different fields are analyzed. By plotting the number of operationalizations instances per category, we will show which categories are predominantly present in one field or the other. Next, by considering each category individually, remarkable (dis)similarities will be identified and presented.

## 5.4. Results

In this section, we will elaborate on the results of our research per research question. First, we will describe the big data analytics and business analytics capabilities we found. Next, we describe the outcomes of the search for learning analytics capabilities. The outcomes are then compared and, finally, combined in the Learning Analytics Capability Model. A dataset with all operationalizations is published online[8].

### 5.4.1. Capabilities for big data analytics and business analytics

Data for this research question was collected in October 2018. Entering our search string in Google Scholar yielded 175 hits. By reading titles and abstracts, it was determined that 150 articles did not meet our inclusion criteria. The remaining 25 articles were combined with the 15 articles already identified by Adrian et al. (Adrian et al., 2018). Removing duplicates left us with 34 unique articles, which in turn were thoroughly read. Based on the inclusion and exclusion criteria, ten studies were marked as key studies (Akter et al., 2016; Brennan et

8 https://www.researchgate.net/publication/339847879_Learning_Analytics_Capability_Model

al., 2018; D. Q. Chen et al., 2015; Cosic et al., 2015; Gupta & George, 2016; Kwon et al., 2014; Ren et al., 2017; Wang et al., 2018; Wang & Byrd, 2017; Wang & Hajli, 2017). In total, the models described in the ten key studies provided 251 different operationalizations. These were coded and based on similarity grouped in 23 different first-order capabilities. The initial coder-agreement was 75%. By categorizing these capabilities based on their characteristics, four second-order capabilities could be distinguished: Data, Management, People, and Technology. We will now elaborate on each of the four second-order capabilities.

#### 5.4.1.1. Data

The category *Data* contains all capabilities related to the use, quality, reporting as well as sourcing and integration of data – see Table 16. In total, this category contains 71 different operationalizations.

| Capability | Description | Operationalization examples |
|---|---|---|
| Data usage | For what goals are big data analytics and business analytics used | Understand trends, scenario planning, predictive modeling |
| Quality | What are the characteristics of data quality | No (input) errors in data, standardization, analytics lead to correct and current information |
| Reporting | How are analytical results presented | Provide actionable insights and proactive recommendations, provide (near) real-time performance metrics |
| Sourcing and integration | What data sources are integrated and how | Data from multiple systems within and outside the organization, integrate in data warehouse |

**Table 16:** Capabilities and operationalization examples for Data

#### 5.4.1.2. Management

With 73 different operationalizations, the category *Management* is the largest of the four second-order capabilities. It involves the benefits of big data analytics and business analytics, governance of analytical processes like capability management, planning and strategy, determining who is responsible and accountable for decisions and their outcomes, benchmarking with external parties, securing funding and investment, as well as the organizational culture and readiness required for the successful deployment of big data analytics and business analytics within an organization – see Table 17.

| Capability | Description | Operationalization examples |
|---|---|---|
| Benefits | What are the benefits of big data analytics and business analytics | Improve the quality of work, lower costs, make work more efficient |
| Capability management | How are organizational capabilities managed | Incorporate analytics into practices, integrate IT leadership and governance infrastructures, ability to reconfigure and leverage capabilities in order to respond to changes |
| Culture and readiness | What are cultural aspects and readiness factors for the adoption of analytics | Make decisions on data rather than instinct, trust in data and tools, encouragement to develop data-driven environment |
| Funding and investment | What kind of funding and investment is necessary and how is it secured | Financial support, given enough time to achieve objectives, consider costs and effects |
| Market | How to align with the external environment | Compare with competitors, customers, and suppliers |
| Performance monitoring | How are the performance of analytical processes and outcomes measured | Clear performance criteria, constantly monitor performance |
| Planning | How to plan the use of analytics in organizational processes | Plan in systematic and formalized ways, enforce adequate plans for analytics introduction, top management create support for analytical initiatives |
| Responsibility and accountability | How are responsibility and accountability managed | Responsibility and accountability are clear, assign decision rights, provide some authoritative autonomy and financial independence |
| Strategy | How to align analytics with organizational strategy | Continuously examine the opportunities the strategic use of analytics, identify important business insights and trends, have a clear vision, have top management promote analytics as a strategic priority |

**Table 17:** Capabilities and operationalization examples for Management

### *5.4.1.3. People*

The third category which can be distinguished from the data is *People*. This capability comprises the (combined) skills and knowledge stakeholders need to have, ways of communicating and collaborating and with whom, as well as the training stakeholders need to receive in order to successfully do their job – see Table 18. In total, this capability is made up of 58 different operationalizations.

| **Capability** | **Description** | **Operationalization examples** |
|---|---|---|
| Collaboration | How is collaboration achieved | Share data and use collaboration portal, coordinate efforts, involve users in planning |
| Combined skills and knowledge | What combined skills and knowledge do people need to have to perform analytics and act on it accordingly | Hold suitable work experience, possess both technical skills and domain knowledge, create and promote a technical innovation team, ability of senior managers and executives to advocate the use of analytics |
| Communication | How will information about analytics will be communicated | Listening carefully to the needs, meet frequently to discuss important issues, share information to have access to all available know-how, eliminate identifiable communications bottlenecks |
| Knowledge | What knowledge do people need to have to perform analytics and act on it accordingly | Business environment, technological trends, critical factors for the success of our organization, exploit existing and explorer new knowledge |
| Skills | What skills do people need to have to perform analytics and act on it accordingly | Learn new technologies, teaching others, entrepreneurial mindset and vision, network, planning and executing work in a collective environment |
| Training | What training do people need to receive | Suitable education, training is provided, staff is well trained |

**Table 18:** Capabilities and operationalization examples for People

### *5.4.1.4. Technology*

The final category relates to *Technology*. This capability concerns the way automation is used in big data analytics and business analytics activities, the role of connectivity, the necessary IT-infrastructure, and the required characteristics of big data analytics and business analytics systems – see Table 19. With 49 operationalizations, this capability is the smallest one.

| Capability | Description | Operationalization examples |
|---|---|---|
| Automation | What is the role of automation in big data analytics and business analytics | Automatic method of maintaining data consistency, automate process for continuously monitoring, automatically notification in case of critical issues |
| Connectivity | In what way can data sources be connected | Data is shared across organization, open system network mechanisms to boost connectivity, cloud-based data warehouse |
| Infra-structure | What infrastructure is necessary for analytics | Visualization tools, databases, analytical interfaces, open-source software, self-service analysis applications, enterprise data infrastructure |
| System characteristics | What are characteristic of (technical) analytical systems | Quick and timely processing, easy to access, adaptable for various analytics tasks, enables work to be shared, protect information |

**Table 19:** Capabilities and operationalization examples for Technology

### *5.4.1.5. Capabilities for big data analytics and business analytics*

When looking at second-order capabilities, it can be noticed that *Technology* is present in all key studies, followed by *Data* which is mentioned in eight of the ten studies. This can hardly come as a surprise, as big data analytics and business analytics are technology-driven and obviously involve the use of data. The management of big data analytics and business analytics and the role of stakeholders are less often present in the existing models. The most frequently mentioned first-order capabilities are *Infrastructure* and *Sourcing & Integration*, which both can be found in eight key studies. Almost all first-order capabilities are present in two or more studies. The only exception is *Training,* which is mentioned in only one study. Moreover, there are just two studies (Cosic et al., 2015; Gupta & George, 2016) in which all second-order capabilities are present.

### 5.4.2. Capabilities for learning analytics

Data for this research question was collected in March 2019. Entering our search string in the ERIC and ACM databases yielded 102 hits. By reading titles and abstracts, it was determined that 90 articles did not meet our inclusion criteria. The remaining 12 articles were combined with the 12 articles already identified by Colvin et al. (2017). Removing duplicates left us with 17 unique articles, which in turn were thoroughly read. Based on the inclusion and exclusion criteria, five studies were marked as key studies – see Table 20. As the research of Colvin et al. (Colvin et al., 2015) is essentially two studies in one – each with their unique objective – we split their research accordingly. This provides us with a total of six key studies that are included in the next phase of our research. From the key studies, 210 operationalizations were extracted and coded according to the *a priori* coding scheme. Initial coder-agreement was 82%, where almost all discrepancies had to do with the classification of first-order capabilities. Disagreements between the two coders were resolved by discussion.



| Reference | Third-order capability | Study objective(s) |
|---|---|---|
| Norris & Baer (2013) | Organizational capacity for student success | Describe the state of the industry and the current and future nature of the analytics gap in higher education. |
| Colvin et al. (2015) (study 1) | Learning analytics readiness factors | Understand how senior institutional leaders perceived learning analytics including the drivers, affordances, and constraints that shape LA within their institutional context |
| Colvin et al. (2015) (study 2) | Dimensions for sustainable uptake of learning analytics | investigating the factors perceived as necessary for establishing sustainable LA implementations that demonstrate long term impact. |
| Ferguson et al. (2014) | ROMA elements | Offer tools and case studies that will support educational institutions in deploying LA at scale to achieve specified learning and teaching objectives. |
| Bichsel (2012) | Analytics maturity factors | Set out to assess the current state of analytics in higher education, outline the challenges and barriers to analytics, and provide a basis for benchmarking progress in analytics. |
| Tsai et al. (2018) | SHEILA elements | Presents a framework that can be used to assist with strategic planning and policy processes for learning analytics. |

**Table 20:** Key studies from the learning analytics domain

Many of the capabilities could be coded according to the *a priori* coding scheme. However, 16 operationalizations do not fit well within this coding scheme. These operationalizations concern privacy aspects and the ethical use of learning analytics. Therefore, we construct a fifth second-order capability: *Privacy & Ethics*. Although it is not mentioned in all studies, this category is present in existing learning analytics models. This is hardly a surprise, as privacy and ethics are often discussed in learning analytics literature (Avella et al., 2016).

#### *5.4.2.1. Privacy & ethics*

The category Privacy & Ethics comprises five different capabilities – see Table 21. They involve the ethical use of learning analytics, the role of human decision-making, the compliance with legal regulations and in particular privacy laws like GDPR, the security of data and information, and transparency about learning analytics.

| Capability | Description | Operationalization examples |
|---|---|---|
| Ethics | How to perform analytics in an ethical way | Policy on ethical use, anticipate ethical dilemmas, establish an ethics committee |
| Human decision-making | What is the role of humans in analytical decision-making | Account for human dimensions, outcomes must be actionable, make no decisions without human evaluation |
| Legal compliance | How to comply with the law | Data ownership, legal frameworks, third party access |
| Security | How to secure data and information | Have information security policies, specify rights and privileges, guarantee data security |
| Transparency | In what way to create transparency about analytics | Be transparent about data use and algorithms, make research reproducible, be clear how ‘success’ is conceived |

**Table 21:** Capabilities and operationalization examples for Privacy & Ethics

#### *5.4.2.2. Capabilities for learning analytics*

Next to an additional second-order capability, the analysis of learning analytics literature also provided some first-order capabilities unique for the learning analytics field – see Table 22.

| Category | Capability | Description | Operationalization examples |
| --- | --- | --- | --- |
| Data | Feedback on analytics | Allows users to provide feedback on the analytics | Provide opportunities to feedback on results, seek feedback, be judged useful by learners |
| Management | Evidence-based and theory-driven | Include evidence and theory in the design of analytics | Blend with proven best practice, be driven by pedagogy, engage with existing literature |
| Management | Implementation and deployment | What factors to consider when implementing and deploying analytics | Integrate in processes, implement top-down, decide on forms of interventions |
| Management | Policies and code of practices | How to (re)-formulate policies | Change written policy, review original policy objectives and vision, consult relevant policies and codes of practice |
| People | Stakeholder engagement | Who to involve in analytics | Engage all stakeholders, involve students, invite teaching staff to contribute |
| People | Stakeholder identification | Who to identify | Identify primary users, senior management, academic teams, internal advocates |

**Table 22:** Capabilities solely present in learning analytics literature



### 5.4.3. Differences and similarities

To identify differences between big data analytics and business analytics capabilities on one hand and learning analytics capabilities on the other, we start with an analysis of the operationalization instances per category. That is, the total number of operationalizations per category. On average, learning analytics key studies provide more operationalizations than studies on big data analytics and business analytics. One of the main reasons for this is the work of Colvin et al. (Colvin et al., 2015), which on its own is responsible for 87 operationalizations. As shown in Table 23, operationalizations for the categories *Data* and *Technology* belong to a large extent to the big data analytics and business analytics literature.

Operationalization of the category *Privacy & Ethics*, on the other hand, can only be found in learning analytics studies. It is remarkable that this category is absent in the big data analytics and business analytics models, even those focusing on healthcare and thus patient data (e.g. Wang & Byrd, 2017). The remaining two categories – *People* and *Management* – are more equally distributed across the literature. However, differences exist also within each category. We now move on to a more in-depth analysis per category.

| **Capability category** | **BDA/BA (n=10)** | **LA (n=6)** |
|---|---|---|
| Data | 71 | 28 |
| Management | 73 | 106 |
| People | 58 | 40 |
| Privacy and ethics | 0 | 16 |
| Technology | 49 | 20 |
| **Total** | **251** | **210** |

**Table 23.** Operationalization instances for big data analytics, business analytics, and learning analytics capabilities

Looking at the category *Data*, one capability is only present in learning analytics literature: *Feedback on analytics*. This capability allows end-users to provide feedback on the (visualization of) analytics they receive. Based on this feedback, analytical outcomes can be improved and better support the beneficial application of insights gained from the analytics. *Sourcing & Integration*, on the other hand, is almost absent from learning analytics models. Nonetheless, it is an important capability as learning analytics ideally uses data from multiple sources (Siemens, 2012) and integration between those sources is paramount for timely and error-free analytics.

With regards to the second-order capability *Management*, it appears that learning analytics models are more internally-oriented than big data analytics and business analytics models, as the latter also considers the external environment (*Market*) they operate in. Learning analytics models, on the other hand, consider evidence and theory, for example, about pedagogy, as important factors for analytical endeavors. Moreover, the learning analytics models mention implementation and deployment as separate capabilities to build to make sure learning analytics integrates with existing processes and considers the appropriate forms of intervention in advance. This is often described in policies and codes of practice, which justifies and elaborates on the use of analytics in educational settings.

In the category *People*, the training of people involved in analytics and the knowledge required for analytics is less often mentioned in learning analytics models than in big data analytics and business analytics models. This is in line with the findings of (Tsai & Gašević, 2017b). However, the identification and engagement of stakeholders is solely mentioned in learning analytics literature.

Looking at *Technology*, the use of automation is only present in big data analytics and business analytics models. Also, connectivity between systems in the organization is hardly mentioned in learning analytics models. This is in line with the previous observation that the sourcing and integration of data sources is underrepresented in learning analytics models. Both academia and practitioners should be aware of these capabilities and consider them when working on learning analytics adoption.

### 5.4.4. The Learning Analytics Capability Model

By researching big data analytics, business analytics as well as learning analytics literature, we found five categories with 34 different capabilities comprising 461 operationalizations. Combining all these capabilities leads to the first version of the Learning Analytics Capability Model: a model specifying what organizational capabilities higher educational institutions need to develop to support the successful adoption of learning analytics and in what way to operationalize them. The model facilitates an increase of learning analytics adoption by higher educational institutions and, as a consequence, helps the field of learning analytics advancing to a higher degree of maturity. We present the Learning Analytics Capability Model in Figure 20.

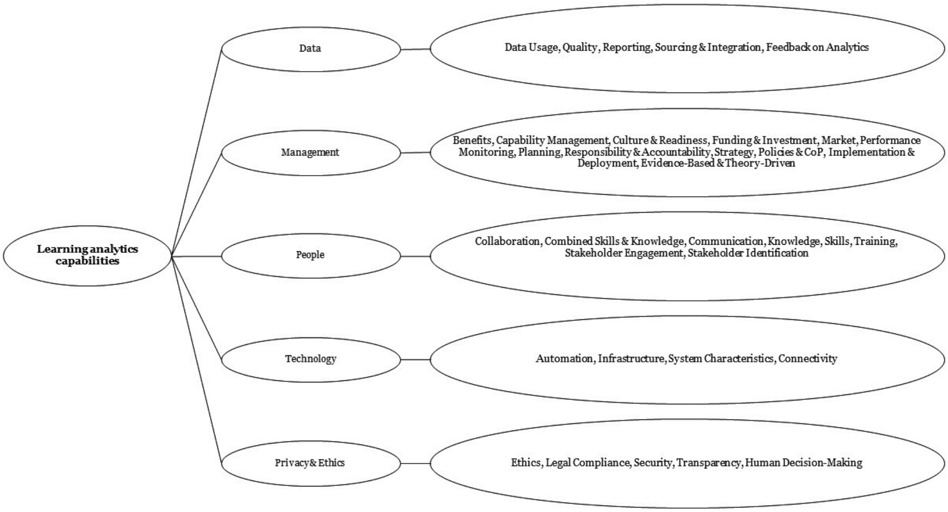


**Figure 20:** Learning Analytics Capability Model

## 5.5. Conclusion and discussion

This literature review provides an answer to the question of what organizational capabilities higher educational institutions need to build for the successful adoption of learning analytics. Because learning analytics is a relatively young research field, we included relevant literature from adjacent research fields, i.e., big data analytics and business analytics. These fields are more mature when it comes to the usage of data to enhance processes and their outcomes. Other research towards learning analytics adoption focuses on certain aspects like policy (e.g., Ferguson et al., 2014; Tsai et al., 2018) or privacy (e.g., Drachsler & Greller, 2016; Hoel et al., 2017; Pardo & Siemens, 2014) but by combining capabilities found in multiple key studies, we now present a model which includes all these aspects: the Learning Analytics Capability Model. Moreover, not only does the model describe the necessary capabilities, it also provides ways to operationalize these capabilities. We thereby enable practitioners, such as senior managers and policymakers, to make strategic and actionable plans towards the adoption of learning analytics in their institution.

The Learning Analytics Capability Model contains five categories: Data, Management, People, Technology, and Privacy & Ethics. These categories comprise 34 different capabilities, for which we provide 461 operationalizations. Some capabilities could only be found in learning analytics literature, for example allowing users to provide feedback on the analysis they receive. However, some other capabilities are presently absent from the learning analytics frameworks we analyzed: i.e., sourcing of data and integration of data sources, the training of stakeholders and learning analytics users in particular, the automation of methods and processes, and connectivity between different systems. We argue that these capabilities must become more prominently present in learning analytics research and practice. When it comes to privacy and ethics, the learning analytics field seems to be quite mature. Although other researchers found that much learning analytics literature does not mention ethical aspects (Viberg et al., 2018), the studies we researched did clearly pay attention to this important aspect. Surprisingly, it is absent in the key studies on big data analytics and business analytics. We recommend researchers and practitioners from these fields to be more aware of privacy and ethics capabilities in the development of big data or business analytics within organizations, and we provide the concrete operationalization of such capabilities extracted from learning analytics literature.

We recognize that our study has limitations. First and foremost, it only relies on secondary data. That is, we conducted a literature review and used existing frameworks to construct our model, so it is not empirically evaluated and validated. Therefore, we consider the current Learning Analytics Capability Model to be the first version and plan to enhance it via a mixed-method approach, i.e., conduct additional case studies to add empirical data to the model and make it more rigorous. Also, the model is yet mainly descriptive and not easily applicable by practitioners who wish to use it. As implementation of learning analytics within an institution is not easy and straightforward, we will enhance the usability of our model for users so it becomes more prescriptive and makes clear how to apply the model to practical settings. A final limitation is the absence of contextual differentiation. All learning analytics-oriented key studies focus on Anglo-Saxon countries, with the Europe-focused SHEILA framework (Tsai et al., 2018) being the only exception. This is in line with the observation of Nouri et al. (Nouri et al., 2019) that at a national or European level, countries yet pay little attention to learning analytics policies and guidelines. As educational ecosystems and thus institutions differ between parts of the world, countries, and even locally, the required capabilities for learning analytics may be different as well. We suggest further research to adapt the Learning Analytics Capability Model for use in specific educational ecosystems to account for differences between, for example, countries.

# CHAPTER

# 6

# 6. Refining the Learning Analytics Capability Model: A Single Case Study[9]

*Learning analytics can help higher educational institutions improve learning. Its adoption, however, is a complex undertaking. The Learning Analytics Capability Model describes what 34 organizational capabilities must be developed to support the successful adoption of learning analytics. This paper described the first iteration to evaluate and refine the current, theoretical model. During a case study, we conducted four semi-structured interviews and collected (internal) documentation at a Dutch university that is mature in the use of student data to improve learning. Based on the empirical data, we merged seven capabilities, renamed three capabilities, and improved the definitions of all others. Six capabilities absent in extant learning analytics models are present at the case organization, implying that they are important to learning analytics adoption. As a result, the new, refined Learning Analytics Capability Model comprises 30 capabilities. Finally, some challenges were identified, showing that even mature organizations still have issues to overcome.*

## 6.1. Introduction

In the past decade, the higher educational domain witnessed the emergence of a new research field: learning analytics. Learning analytics is the analysis and visualization of learner data with the goal to improve learning and the learning environment (LAK, 2011). One of its main drivers is IS/IT, as the digitalization of education led to the increased availability of learner data (Ferguson, 2012). However, despite the promising results, the uptake of learning analytics by higher educational institutions remains low (Gašević et al., 2019). One of the main causes for this is the complexity of implementation, which requires attention to many different dimensions. Several learning analytics adoption models are designed, for example, ROMA (Ferguson et al., 2014), SHEILA (Tsai et al., 2018) and LALA (Pérez-Sanagustín et al., 2019). However, these existing models often focus only on specific elements like policy or privacy and ethics, or lack descriptions on how to operationalize important dimensions. These shortcomings limit the practicality of the models and help higher educational institutions only to a certain degree. To

9 This work was originally published as: Knobbout, J.H.; van der Stappen, E.J.; and Versendaal, J.M. (2020) Refining the Learning Analytics Capability Model: A Single Case Study. *AMCIS 2020 Proceedings*. 9. https://aisel.aisnet.org/amcis2020/is_education/is_education/9

overcome the shortcomings and support higher educational institutions in their quest to adopt learning analytics successfully, the Learning Analytics Capability Model is designed (Knobbout & van der Stappen, 2020a). As higher educational institutions need to make strategic decisions about resources and institutional capacities (Arnold et al., 2014), the model takes a resource-based perspective. The resource-based view attributes organizational performance to its resources and is used to study the uptake of business analytics and big data analytics - two research fields adjacent to learning analytics. The Learning Analytics Capability Model describes what organizational capabilities higher educational institutions need to build to support the uptake of learning analytics. Moreover, the model provides clear operationalizations of these capabilities, helping institutions' senior management and policymakers to implement learning analytics successfully. It is the result of a literature review towards capabilities for business analytics, big data analytics, and learning analytics. It comprises 34 different capabilities divided over five second-order categories.



Momentarily, the Learning Analytics Capability Model is only grounded in theory. Drawing from Design Science Research, the evaluation and refinement of a made artifact (the model) is an important part of the design process (Hevner, 2007). Therefore, before the Learning Analytics Capability Model can be used in practice, it needs thorough evaluation and refinement. To perform the first refinement of the model, we conduct a single case study. This way, we include empirical data and practical experience to the model. This paper provides an answer to the research question "*How can the Learning Analytics Capability Model be evaluated and refined based on empirical data from a single Dutch higher educational institution that is mature in the use of learner data to improve learning*?" The case study is conducted at a higher educational institution that is mature in the use of data to improve education and comprised four interviews with different stakeholders. Transcriptions of the interviews are coded and compared with the theoretical Learning Analytics Capability Model. In turn, the model is improved as the interviews provide new insights into the capabilities necessary for successful learning analytics implementation.

The remainder of this paper is structured as follows. First, we describe relevant work related to the resource-based view and capabilities for big data analytics, business analytics and learning analytics, as well as the Learning Analytics Capability Model. Next, we elaborate on the research method, including case selection and interview procedures. We then comparison of the theoretical Learning Analytics Capability Model with the collected, empirical data, and describe the refinements we make to the model. Next, we draw conclusions from the results and answer the research question. Finally, we discuss our work and provide directions for future research.

## 6.2. Theoretical background

The resource-based view attributes organizational performance to its resources and, in order to obtain sustained competitive advantages, these must be valuable, rare, inimitable, and non-substitutable (Barney, 1991; Bharadwaj, 2000). Different kinds of resources can be distinguished: financial resources, physical resources, human resources, technological resources, organizational resources, and reputation (Barney, 1991; Grant, 1991). Moreover, resources can be subdivided into two distinct groups: assets and capabilities (Helfat & Peteraf, 2003; Wade & Hulland, 2004). Assets involve anything which can be deployed by an organization to create, produce and offer its goods or services to a market and can be either be tangible, intangible or personnel-based (Bharadwaj, 2000). In contrast, capabilities are repeatable patterns of actions in the use of these assets (Wade & Hulland, 2004) and involve "complex patterns of coordination between people and between people and other resources" (Grant 1991, p. 122). Capabilities refer to an organization's capacity to deploy other resources and ownership cannot be transferred between organizations as they are deeply embedded in the organization (Makadok, 2001). As a result, capabilities are no commodities that can be bought, but they need to be built to effectively interact with the organizational processes and procedures.

The resource-based view is utilized to study the capabilities needed for big data analytics (Gupta & George, 2016). Big data analytics provides the required knowledge about the handling, analysis, and visualization technologies of large and complex datasets (H. Chen et al., 2012). Based on the analysis of 15 key studies on big data analytics, Adrian, Abdullah, Atan and Jusoh (2018) show that possessing the right capabilities is an important factor for organizations that adopt big data analytics. These capabilities relate to, among others, management, technology, talent, and information processing. The resource-based view is also applied to study capabilities for business analytics (Cosic et al., 2015). Business analytics uses the analysis of data to understand and manage businesses more effectively (Kohavi et al., 2002). Business analytics and big data analytics are comparable to analytics in educational settings (Barneveld et al., 2012; Picciano, 2014). Like its non-educational counterparts, learning analytics can bring competitive advantages to the educational domain when institutions invest in resources and institutional capacities (Arnold, Lynch, et al., 2014).

Learning analytics is “the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environment in which it occurs” (LAK, 2011). The learning analytics process involves multiple steps: the generation and collection of learner data, the analysis and visualization of these data, and interventions (Clow, 2012).

Learning analytics could affect and improve learning processes, learning environments, student performance, and departmental performance (Knobbout and van der Stappen, 2018). Many higher educational institutions are interested in learning analytics, but not many have already adopted it to address institutional and educational challenges (Gašević et al., 2019). Issues related to implementation may be technical but also involve strategic planning and policy (Tsai & Gašević, 2017b). Empirical research by Ifenthaler and Yau (2019) shows that stakeholders often can identify the resources necessary for learning analytics adoption but that most institutions still need to build and attain these required resources. However, despite the importance of having the right resources and capabilities, the resource-based view is only recently used to study capabilities for learning analytics (Knobbout & van der Stappen, 2020a). Based on existing literature on business analytics, big data analytics and learning analytics, the authors designed the first, theoretical version of the Learning Analytics Capability Model. Via the open coding of 461 operationalizations found in 15 key studies, 34 different capabilities were identified. These capabilities could then be categorized into five second-order capabilities: *Data, Management, People, Technology*, and *Privacy & Ethics*. Six capabilities were only present in literature on business analytics and big data analytics: *Sourcing & Integration, Market, Knowledge, Training, Automation*, and *Connectivity*. Since these research fields interconnect with learning analytics, these capabilities were adopted to the Learning Analytics Capability Model – a process called exaptation (Gregor & Hevner, 2013).

To enhance its rigor, the Learning Analytics Capability Model must be evaluated and refined (Hevner 2007; Hevner et al. 2004). Evaluation is about identifying weaknesses in the designed artifact. In our situation, we must research whether any capabilities are missing in our model, whether the exapted capabilities are indeed present at the case organization, and whether the capability definitions are explicitly enough to capture the ways capabilities are operationalized. Based on the outcomes of this identification, the model will be refined. To perform the task of evaluation and refinement, we conduct a case study at a higher educational institution that is experienced in the use of learner data to enhance learning. The research methodology for this study is described in the next section.

## 6.3. Research method: protocol definition and execution

In our study, we opt for a case study to verify and complement what has been found from literature. Using the case study, we study contemporary events without the need to control behavioral events (Yin, 2013). A case study is a suitable evaluation method to study a designed artifact in an organizational environment (Hevner et al., 2004). For preparation, we take notice of the case study protocol guidelines from Maimbo and Pervan (2005).

### 6.3.1. Preamble

For the case study, we take interviews and consult documentation. We ask the interviewees to agree on the proposed anonymous scientific publication of the results of the case study. We explain the reason for the research to the interviewees and ask them to sign a confidentiality agreement. After transcribing the interviews, the audio files are deleted. Data is stored in compliance with the GDPR.

### 6.3.2. Procedure

Especially in Europe, institutions that successfully apply learning analytics are scarce (Gašević et al., 2019), thus providing an unusual case. This justifies the use of a single case study that focuses on a single unit of analysis (Yin, 2013). The unit of analysis in our study is the analytics team and its internal customers of a Dutch university that uses learning analytics across the organization. We select the case by consulting a group of learning analytics experts in the Netherlands. They suggest a particular case organization - which we anonymize to The Netherlands University (TNU) - as it was the first higher educational institution in the Netherlands publishing a code of practice related to the analysis of student data and because it already has a couple of years' experience with organizational-broad use of learning analytics. TNU is a large academic university with around 25,000 enrolled students and located in the Netherlands. To ensure the organization's learning analytics maturity, we additionally apply the EDUCAUSE maturity model (Bichsel, 2012). This model is partly based on the work of Davenport and Harris (2007) but adjusted to the educational domain and can be used to score various dimensions important to learning analytics uptake. TNU scores well on each dimension of the EDUCAUSE model, that is, four on a five-point scale. Such a score is comparable to stage 4 (analytical company) of the Davenport and Harris model, so the organization is suitable for the study at hand. The interviews took place at TNU in December 2018 and January 2019, and the interview duration was between 40 and 90 minutes.

Learning analytics is a multidisciplinary field including “educators, learning scientists, computer scientists, administrators, and policymakers” (Suthers & Verbert, 2013), and consequently, the interviews need to reflect this. To ensure a broad view of the topic, we select interviewees with different roles within the learning analytics process at TNU, including users of learning analytics within the organization. The first person to interview is the manager of the analytics team, with whom contact was already established. From there on, we apply a snowballing technique to select the next interviewees. To further enhance the quality of our study, we also request relevant (internal) documentation to support statements made during the interviews. This form of data triangulation increases the quality of the study (Yin, 2013). In total, four interviews are conducted. We send leading questions from the interview protocol (in Dutch) in advance so the interviewees can prepare their answers and bring relevant material to the interview. Interview questions are derived from the theoretical Learning Analytics Capability Model (Knobbout & van der Stappen, 2020a)[10]. The characteristics of each interviewee and on what category of the theoretical model the interview questions were focused, are described in Table 24.

6

| | **Job title** | **Relation to learning analytics** | **Question categories** |
|---|---|---|---|
| Interviewee A | Data engineer | Member of the analytics team | Data, Technology, Privacy & Ethics |
| Interviewee B | Project leader | Manager of the analytics team | Data, Management, People, Technology, Privacy & Ethics |
| Interviewee C | Student advisor | User | Data, Management, People, Privacy & Ethics |
| Interviewee D | Policymaker | User | Data, Management, People, Privacy & Ethics |

**Table 24.** Interviewee Characteristics

10 Interview questions (in Dutch) are available on request by contacting the first author.

### 6.3.3. Research instrument

We use open semi-structured face-to-face interviews to collect data. We create an interview protocol with procedures and questions and discuss this with a panel of researchers. We anticipate fine-tuning this protocol after each interview. Interviewees are requested to bring relevant archival data – documents, presentations etc. - to the interviews. Also, publicly available archival data (the code of practice, presentations) from TNU are collected. The use of these data is twofold: 1) to help guide the interviews and allow for ad-hoc questions about the documentation brought by the interviewees, and 2) to later clarify and verify statements made by the interviewees.

### 6.3.4. Data analysis guidelines

We record the interviews and transcribe each interview verbally. In line with suggestions by Runeson and Höst (2009), the main researcher performs the transcription, while multiple researchers do the subsequent coding of the transcriptions. Due to technical malfunctions, only half of the interview with interviewee D was recorded. Luckily, notes were taken during the interview, and these notes were used to reconstruct the interview. We send transcriptions to the interviewees so they can check for errors or misinterpretations. No objections or requests for change were received. Next, we will code the transcriptions in Atlas. ti. We apply the principle of axial coding, where the codes are structured based on existing knowledge (Strauss and Corbin 1990). A-priori coding comes from our initial Learning Analytics Capability Model (Knobbout & van der Stappen, 2020a). We place interview fragments not matching the existing codes in a separate category for later analysis. Archival data is used to clarify and validate comments made by the interviewees. Two researchers work in parallel and code the same interview. The results are then to be discussed, after which the coding protocol with code definitions can be adjusted. This process is to be repeated until all interviews are coded. After the coding process is finished, we discuss the results with a third researcher. The interviews are in Dutch so all quotes in this paper are our translations. In the next section, we present the results of our analysis.

## 6.4. Findings

The analysis of the transcriptions of four different interviews at the case organization resulted in 424 assigned codes. Next to the five categories of capabilities, a list of challenges emerged from the data – see Table 25. The interviews of interviewees A and B contained significant more codes than the interviews of interviewees C and D. This is not surprising, as A and B work at the back end of the analytics process while C and D are primarily users and only see the outcomes. Codes related to the category *Data* appear the most, followed by *People*. Codes about *Privacy & Ethics* appeared the fewest, and this is the only category that is not mentioned in all interviews. Collected documentation included the organization's policy on learning analytics, a list with (learning) data sources and presentations about the subject given by members of the analytics team. The most important conclusion is that all capabilities necessary for learning analytics adoption are present in the current Learning Analytics Capability Model. Nonetheless, several improvements could be made to the model.

| | Interviewee | | | | |
|---|---|---|---|---|---|
| **Category** | **A** | **B** | **C** | **D** | **Total** |
| Challenge | 18 | 9 | 7 | 9 | 43 |
| Data | 49 | 37 | 18 | 14 | 118 |
| Management | 31 | 48 | 8 | 11 | 98 |
| People | 27 | 46 | 12 | 16 | 101 |
| Privacy & Ethics | 3 | 12 | 0 | 3 | 18 |
| Technology | 30 | 11 | 4 | 1 | 46 |
| **Total** | **158** | **163** | **49** | **54** | **424** |

**Table 25.** Codes Assigned to Each Interview



### 6.4.1. Comparison with the theoretical Learning Analytics Capability Model

When we compare the theoretical model with the collected data, not all capabilities from theory appear in practice at TNU. The capabilities *Performance Monitoring* and *Human Decision-Making* were not mentioned in any of the interviews. *Performance Monitoring* relates to the monitoring of the learning analytics process performance, i.e., does the analytics improve learning according to pre-defined criteria. One reason for the interviewees not mentioning this aspect could be because the actual interventions are performed by other stakeholders who do not relay the improvements back to the analytics team. However, to enhance the analytical process, it would be good for TNU to keep track of the improvements made on education. The absence of *Human Decision-Making* in the interviews might be explained by the fact that in the Netherlands, fully automated decision-making based on student data is prohibited by law (Engelfriet et al., 2017) so humans are by default involved in the decision-making process.

In a previous study (Knobbout & van der Stappen, 2020a), we learned that several capabilities, e.g., *Sourcing & Integration* and *Automation*, were present in the literature on business analytics and big data analytics but absent from existing learning analytics models. Based on the principle of exaptation (Gregor & Hevner, 2013), they were included in the Learning Analytics Capability Model but their relevance to the learning analytics domain remained unclear. During the interviews, however, these absent capabilities were mentioned by the interviewees and thus appear to be necessary for the successful uptake of learning analytics within an institution. For example, documentation shows 52 different data sources, so *Sourcing & Integration* is an important capability for TNU. Data could be extracted from the Management Information System, the Virtual Learning Environment, the Enterprise Resource system, and many more. Interviewee A comments on the integration of different sources: "*it is easy to do because we use linking tables, so it all translates to each other*". The capability *Automation* plays a role there as well: "*We have automated as much as possible so the [analytics] team members are not wasting time linking things together*" (Interviewee B). Also, by automating processes, the quality of data is secured: "*After each step, automated tests are conducted […] so we can check each step*" (Interviewee B).

The capability *Training* is clearly present at TNU: "*Once, we had a statistics course with the whole team for a full week*" (Interviewee A) and "*with each other, we procure education. [The team members] decide what skills we need to develop*" (Interviewee B). Not only the analytical team receives training, but users do as well: "*they got a short training - one afternoon - on how to interpret [the analytical outcomes]*" (Interviewee B). This links to the capability *Knowledge,* as both the team members and the users often need extra training to attain relevant knowledge: "*At the moment, we don't have much knowledge about text mining*" (Interviewee A), *and* "*you really need to know what you're looking at*" (Interviewee C). Especially users should have the right knowledge about how to interpret the analytical outcomes, as they are the ones performing interventions.

Higher educational institutions do not operate in a vacuum. That is, some comments were made with regards to the capability *Market*: "*another university is interested in procuring [our method] from us*" (Interviewee B) and "*They used material from England, from [JISC]. Well, we used the things we thought to be important*" (Interviewee B). An adjacent capability is *Connectivity*, which describes the external connection between systems: "*We use SAP, but other institutions use other systems so transferring data is difficult*" (Interviewee D).

The presence of the capabilities unique for business analytics and big data analytics at TNU confirms their adoption to the learning analytics domain and thus their inclusion in the Learning Analytics Capability Model.

### 6.4.2. Refinements to the theoretical Learning Analytics Capability Model

An important outcome of our study is that no capabilities seem to be missing to the Learning Analytics Capability Model. Nonetheless, several refinements[11] could be made based on the collected data. Refinements concern the merging of capabilities, the renaming of capabilities, and/or the reformulation of the capability definitions. Most changes affect the categories *Data, Management,* and *People*. In practice, some capabilities overlap to such a degree that merging them would improve the model. Also, refinement was deemed necessary when the capability definitions from the theoretical model did not reflect the statements made by the interviewees. Often, the name or the definitions needed clarification to describe better what exactly is covered by each capability. Refinements were established through multiple rounds of discussion between three researchers (two of whom coded the data). During our analysis, some challenges were identified. These could often be linked to certain capabilities, but they do not have a solution yet. We present some of the found challenges to guide future research. We will now discuss the most important changes per type of refinement.



### 6.4.3. Merging capabilities

In theory, the difference between knowledge and skills is easily described. In practice, however, it is hard to distinguish between skills without considering the need for related knowledge and vice versa. For example, Interviewee B describes a skill needed by members of the analytics team: *"programming in R"*. Although this quote only implies the need for a certain skill, without knowledge about both programming and R, this skill is impossible to master. This is often the case. Therefore, we merge capabilities *Knowledge* and *Skills* in the already existing capability *Combined Skills & Knowledge*.

The same principle applies to the capabilities *Stakeholder Identification* and *Stakeholder Engagement*. Without the ability to identify the right stakeholders, they cannot be engaged. Interviewee C comments: *"Academic advisors are of course involved"*. These advisors need to be identified before they can be involved

11 See https://hbo-kennisbank.nl/details/sharekit_hu:oai:surfsharekit.nl:88cdbc33-c3d0-4748-b81d-263f6ad44876?q=LACM for an overview of the original and the refined capabilities, including their definitions.

in the analytical process. As a result, we merge the *Stakeholder Identification* and *Stakeholder Engagement* into the new capability *Stakeholder Identification & Engagement*.

The capability *Planning* describes the planning of the learning analytics implementation. However, the Learning Analytics Capability Model also comprises a separate capability *Implementation & Deployment* that describes the actual implementation. Although *Planning* relates to the process before and *Implementation & Deployment* to the process during implementation, these two capabilities overlap. Moreover, after implementing learning analytics into the organization, the work is not finished. Learning analytics processes need to be refined, optimized, and planned. As stated by Interviewee B: "*We have a long list of things we want next – new datasets, improving quality of tests, doing things we believe are interesting for TNU.*" Hence, we both combine and rename these two capabilities to *Implementation Deployment & Application*.

### 6.4.4. Renaming capabilities

The capability *Market* is renamed to *External Environment*, as not all external collaboration is commercial as the term 'Market' suggests. As mentioned earlier in this paper, this capability describes all influences from outside the organization. This includes the use of material and tools from external parties, the hiring of external personnel, requests and demands from external (governmental) bodies, and sharing materials, knowledge and experiences with other (higher) educational institutions.

*Benefits* describes the benefits of learning analytics for the organization, which is not a capability per se. It is important though to identify the benefits to education as this should be part of the learning analytics design (Wise, 2014). Higher educational institutions, therefore, should be able to describe the benefits they want to achieve with learning analytics. In the Learning Analytics Capability Model, we rename the capability *Benefits* to *Identifying Benefits*.

*Capability Management* is needed to manage existing capabilities. Also, it should include the development and reconfiguration of these existing capabilities. This principle aligns with the concept of dynamic capabilities: "routines by which firms achieve new resource configuration" (Eisenhardt & Martin, 2000). To better reflect the ongoing change of capabilities already present at the organization, we rename this capability to *Capability Development*. Moreover, we use this capability to describe challenges for which the solutions are already known, i.e., the organization knows how to deal with problems.

### 6.4.5. Improving capability definitions

Besides many small changes to the capability definitions, four major improvements were made. First, *Data Usage* relates to the goals of analytical processes. However, during the interviews, multiple comments about intervention strategies were made, for example, improved communication towards students (Interviewee B, Interviewee D), or supporting students with data-informed planning tools (Interviewee B). The attention for interventions is not surprising, as it is part of the learning analytics process (Clow, 2012). Also, at TNU data is often aggregated: "*We aggregate to a level on which we can do analyses*" (Interviewee A). Aggregating data is a form of summarizing and thus use of data. So, Data Usage is broader than only the goals of analytics, and we change its definition to "in what way data analysis is used to improve education. It contains data aggregation, different kinds of analysis, the goals of the analysis, and the interventions performed based on the outcomes of the analysis."



Next, three capabilities need better definitions as they partly overlap and caused confusion while coding the interview data: *Reporting, Communication,* and *Collaboration*. *Reporting* describes in what way the outcomes of data analysis are presented to the various stakeholders (dashboards, reports, etc.) and what the requirements of the presentation are. In contrast to *Communication*, the flow of information is one-way, i.e., from the party delivering the analytical outcomes to the party receiving them. That said, *Communication* relates to the flow of information between (groups of) stakeholders. This includes communication between users and the party delivering the learning analytics about the needs and possibilities ('demand and supply'), the communication mechanisms and the types of information that are shared between different (groups of) stakeholders. Finally, *Collaboration* describes the active cooperation between parties - either within a group of stakeholders or between groups of different stakeholders, and either internal or external. This capability also includes the mechanisms via which collaboration is achieved.

### 6.4.6. Challenges mentioned by interviewees

The case study provided insight into the capabilities present at TNU but also in some of the challenges faced by the organization when maturing in the use of learner data. These challenges often relate to existing capabilities, like *Sourcing & Integration: "sometimes we need to couple [data sets] ourselves because others* *did not couple or aggregated it in the right way"* (Interviewee A), or *Training*: "*Instructional video's or manuals could help, but they are outdated pretty fast because the developments go very quick"* (Interviewee D). Solutions are not always clearly present, making it hard to mature in the use of learning analytics. We take the challenges into consideration and use them as directions for future research.

## 6.5. Conclusion

The aim of this study is to provide an answer to the question "*How can the Learning Analytics Capability Model be evaluated and refined based on empirical data from a single Dutch higher educational institution that is mature in the use of learner data to improve learning*?" By conducting a single case study, we collected empirical data on capabilities important to the successful uptake of learning analytics. We take a resource-based perspective, and because capabilities cannot be transferred between organizations, we emphasize the importance of higher educational institutions to develop their own capabilities. As Interviewee D correctly mentions: "*even when we sell our model or [the manager of the analytics team], another organization cannot do what we do*". Based on the analysis of the collected data, seven capabilities from the previous Learning Analytics Capability Model are merged, three of them are renamed, and all definitions are improved. Eventually, we made significant improvements to the model. The new, refined Learning Analytics Capability Model comprises 30 capabilities – see Figure 21.

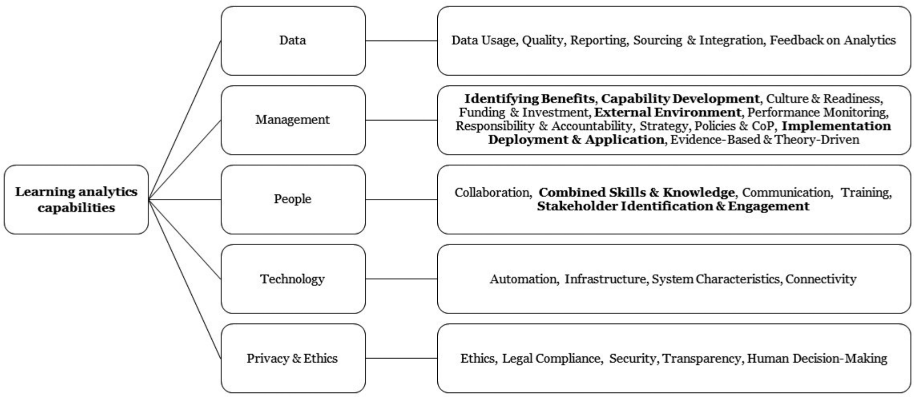


**Figure 21.** Refined Learning Analytics Capability Model (Changed Capabilities in Bold)

Our study has both scientific and practical relevance. First, the Learning Analytics Capability Model provides a relevant and rigorous set of clearly defined learning analytics capabilities. Second, to our knowledge, the Learning Analytics Capability

Model is the first model using the resource-based view in relation to learning analytics. This helps academics to extend their vision on learning analytics adoption as it eases the comparison between analytical capabilities of higher educational institutions and those of organizations outside the educational domain. Third, the Learning Analytics Capability Model identified some important capabilities missing from extant learning analytics models, helping researchers to improve these models. Finally, the model helps practitioners to develop the right capabilities so higher educational institutions can adopt learning analytics more efficiently.

## 6.6. Discussion & future work

We are aware that our work has some limitations. First, we only interviewed four stakeholders at a single organization. Although care was taken that the interviewees have different roles, experiences, and insights, it is possible that more interviews would have brought up new information important to our study. The same applies to the selection of a single organization – other organizations might have different capabilities yet unknown to us and therefore absent from our model. This study, however, is the first evaluation of the Learning Analytics Capability Model and a multiple-case study with a broad selection of different organizations is planned in the coming year. The refined model is currently used as blueprint to develop a digital tool that allows institutions to measure what learning analytics capabilities they already possess and what others they still need to build. Data collected via this tool will be used to further refine our model. This follows the call of Hevner (2007) for multiple iterations of the design cycle. Second, in addition to interviews, archival data were used for clarification and verification purposes. These data, however, were not coded themselves. Therefore, relevant information in the collected documentation might be missed. We suggest the coding of these data and comparison with the results from the study at hand as a future research activity. During the analysis, we discovered of some challenges currently faced by the case organization. Solutions to these challenges are not always easy. They might be widespread ("[the lack of] *statistical knowledge is often a problem*", Interviewee D), hard to solve ("*in most BI-systems, field names are restricted*", Interviewee A), or take considerable time and effort ("*you need to get the managing directors on board*", Interviewee B). In our future research, we will focus on these challenges, helping higher educational institutions - and the educational domain in general - to mature in the adoption and use of learning analytics.

# PART VI

Evaluation

# CHAPTER

# 7

# 7. A Comprehensive Model to Support the Adoption of Learning Analytics: A Mixed-Method Approach[12]

*Although learning analytics benefit learning, its uptake by higher educational institutions remains low. Implementing learning analytics is a complex undertaking and higher educational institutions lack insight into what organizational capabilities must be developed before learning analytics can be successfully used institution-wide. To address this problem, a capability model for learning analytics was developed. It intends to support practitioners such as program managers, policymakers, and senior management by providing a comprehensive overview of necessary capabilities and insight into operationalizing these capabilities. The model is grounded in the resource-based view. This paper describes its ex-post evaluation via a mixed-method approach. Qualitative data is collected during pluralistic walk-throughs with 26 participants at five educational institutions in the Netherlands and Belgium and a group discussion with seven learning analytics experts. Quantitative data about the model's perceived usefulness and ease-of-use is collected via a survey (n = 23). Our study shows that the model is positively evaluated by the participants. Hence, the model is concluded useful for planning the implementation of learning analytics, is perceived useful by practitioners, and contains all necessary elements. Moreover, this study shows the applicability of pluralistic walk-throughs as method for ex-post evaluation of IS artifacts.*

## 7.1. Introduction

The digitization of education led to the increased availability of learner data (Ferguson, 2012). In turn, these learner data could be collected, analyzed, and used to perform interventions to improve education (Clow, 2012). This process is called *learning analytics (LA)*. Being highly dependable on the swift and correct handling and analysis of data, Information Systems (IS) play a paramount role in LA processes (Dawson et al., 2019; Ferguson, 2012; Nguyen et al., 2020; Rubel & Jones, 2016). LA are effective in improving student outcomes (Foster & Francis, 2019). For that reason, higher educational institutions (HEIs) try to adopt LA to their educational processes. Over the years, researchers designed models,

12 This work was originally published as: Knobbout, J.H.; van der Stappen, E.J.; Versendaal, J.M.; and van de Wetering, R. (2021) A Comprehensive Model to Support the Adoption of Learning Analytics: A Mixed-Method Approach. *Manuscript submitted for review.*

frameworks, and other instruments to support LA' uptake (Arnold, Lynch, et al., 2014; Bichsel, 2012; Ferguson et al., 2014; Greller & Drachsler, 2012; Siemens et al., 2013; Tsai et al., 2018). However, despite these models' existence, only a few institutions successfully implemented LA at scale (Dawson et al., 2018). Extant implementation models such as the Learning Analytics Framework (Greller & Drachsler, 2012) and the Learning Analytics Readiness Instrument (LARI)(Arnold, Lonn, et al., 2014) are too generic and do not offer practical insights into operationalizing critical dimensions for LA adoption (Broos et al., 2020). Other models like the Supporting Higher Education to Integrate Learning Analytics (SHEILA) framework (Tsai et al., 2018) and the Rapid Outcome Mapping Approach (ROMA) (Ferguson et al., 2014) only focus on specific elements like policy-development and do not provide solutions to challenges identified by institutions while using the models (Tsai et al., 2018). A capability model for LA (Figure 22) was developed to overcome this knowledge gap (Knobbout et al., 2020; Knobbout & van der Stappen, 2020a).

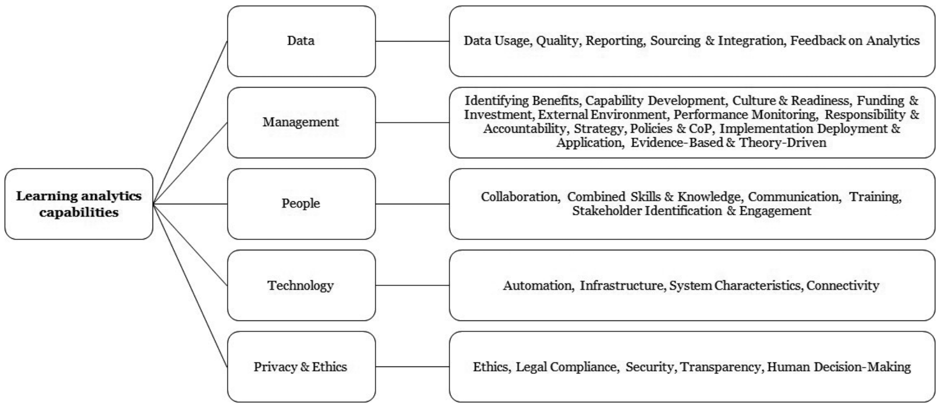


**Figure 22:** Capability Model for LA.

As the lack of the right organizational resources is a major reason for HEIs not to adopt learning analytics (Tsai & Gašević, 2017b), the capability model takes a resource-based perspective. The resource-based view (RBV) provides a theoretical basis for improving processes, performance and creating competitive advantages via the organization's resources (Barney, 1991; Grant, 1991). Resource-based capabilities describe how physical, human, and organizational resources should interact to achieve these benefits (Amit & Schoemaker, 1993; Makadok, 2001). Although capability models exist in adjacent research fields (e.g., (Cosic et al., 2015; Gupta & George, 2016), no such model is available for the LA domain. The capability model evaluated in this study is designed through the application of the Information Systems Design Science Research (IS DSR) framework (Hevner et al., 2004). IS DSR can be applied to design various kinds of IS/IT artifacts, including scientific models. First, in a study that can be classified as exaptation research (Gregor & Hevner, 2013), existing knowledge on capabilities important for the successful uptake of analytics are exapted from business analytics and big data analytics literature to the LA domain (Knobbout & van der Stappen, 2020a). This led to a theoretical capability model. Subsequently, the model was refined by conducting a case study, thus including empirical data in the model (Knobbout et al., 2020). The refined model comprises 30 different organizational capabilities, divided over five categories. Similar to Greller & Drachsler (2012), we argue that each category must be instantiated to achieve successful LA at an institution. To increase the practical relevance of the model, multiple ways to operationalize each capability are included. This way, users of the model do not only get insight into what capabilities they need to (further) develop at their institutions but also how to proceed towards this aim. The model answers the questions `what capabilities do we need?' as well as `in what way can we build them?' Intended users of the model are practitioners such as program managers, policymakers, and senior management who need to plan the implementation of LA at their institution. The model focuses explicitly on Dutch HEIs. Although LA is increasingly used in the Netherlands (Nouri et al., 2019), Dutch HEIs still face challenges related to data integration, ownership of data and analyses, sound infrastructure for learner data, and privacy (van der Spek, 2018; Vereniging van Universiteiten et al., 2017). Most research towards LA implementation is located in the UK, USA, and Australia (Yau & Ifenthaler, 2020), the SHEILA framework the only exception (Tsai et al., 2018). As national educational systems and culture differ, insights from extant models are for Dutch HEIs only useful to a certain degree. To support these institutions in adopting LA at scale, the capability model aims at this specific educational context.

An essential step in design science research is the evaluation of the designed artifact (Hevner et al., 2004; Peffers et al., 2012; Prat et al., 2015; Venable, 2010). After its design, formative evaluation, and refinement, the model's validity must now be demonstrated by returning it "into the environment for study and evaluation in the application domain" (Hevner, 2007). Since the artifact is already developed and will now be evaluated with real users in a real context, we apply an ex-post naturalistic strategy (Pries-Heje et al., 2008). Our paper describes this ex-post evaluation of the capability model in the context of Dutch-speaking educational institutions. The relevance of our work is threefold. First, it provides an empirically validated capability model that is 1) effective in its task of supporting practitioners when implementing LA at their institution, 2) perceived useful by its users, and 3) complete in the sense that it contains all necessary elements. Our research shows what capabilities are often overlooked when planning the implementation of LA. It also highlights the need to remember that LA is about learning and, as a consequence, educational elements should be well represented in every LA implementation model. Second, it further established pluralistic walk-throughs (Bias, 1994) as evaluation method in IS DSR. Most often, pluralistic walk-throughs are applied to research user interfaces' usability, but in line with other scholars (Dahlberg, 2003; Emaus et al., 2010; Kusters & Versendaal, 2013), we used pluralistic walk-throughs to evaluate the capability model. Our work substantiates pluralistic walk-throughs' applicability as an evaluation method in situations where artifacts need to be used in real-world settings but time is constrained. Third, the use of design science research in the LA domain is limited and, to the best of our knowledge, only done recently (Dawson et al., 2019; Knobbout & van der Stappen, 2020a; Nguyen et al., 2020). We add to the design science research knowledge base by applying design science principles to design and evaluate an artifact for the LA domain.

The structure of this paper is as follows. In the next section, we will provide an overview of our study's background and relevant literature. We then present the methodology we used for our research, i.e., a mixed-method approach by conducting pluralistic walk-throughs, a group discussion with experts, and a survey. Next, we describe the analyses we performed and their outcomes. Finally, we conclude our work with a discussion on the results and their implications, the limitations of our study, and directions for future research.



## 7.2. Background

In this section, we describe the background of our study based on relevant literature. We start with a description of existing LA models and frameworks. Next, the design process of the capability model is described. Finally, we describe the evaluation's role in the design process from an IS DSR perspective.

### 7.2.1. Existing models supporting learning analytics adoption

LA is "the measurement, collection, analysis, and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environment in which it occurs" (LAK, 2011). Although LA can enhance various aspects of education, not many HEIs' adopted LA at scale yet (Dawson et al., 2018; Gašević et al., 2019). A systematic literature review by Viberg et al. (2018)

shows that there is little evidence that LA are widely used. They conclude that LA tools are deployed in only 6% of the analyzed research papers. Similar findings are reported by (Tsai & Gašević, 2017a), who interviewed LA experts and surveyed European HEIs. Only 2 out of 46 (4%) surveyed institutions achieved institution-wide LA. Challenges for implementing LA can relate to technology, the shortage of leadership capabilities, or insufficient training opportunities for end-users (Tsai & Gašević, 2017b). Several models to overcome implementation challenges and support the uptake of LA exist. Scholars distinguish three kinds of models: input, output, and process models (Broos et al., 2020; Colvin et al., 2017; Dawson et al., 2018). Input models such as the Learning Analytics Framework (Greller & Drachsler, 2012) and the LARI (Arnold, Lonn, et al., 2014) describe what dimensions are required for LA adoption. Output models such as the Learning Analytics Sophistication Model (Siemens et al., 2013) describe important dimensions and what outcomes to expect from it and how to grow towards the desired outcomes. Thirdly, process models like the ROMA (Ferguson et al., 2014) and the SHEILA framework (Tsai et al., 2018) map a sequence of processes to achieve LA adoption. However, all these models have their shortcomings. They are often not grounded in management theory, they focus on only one aspect of implementation like policy-development, or they lack clear operational descriptions on developing essential dimensions. To overcome these gaps, a capability model for LA is designed (Knobbout et al., 2020; Knobbout & van der Stappen, 2020a).

### 7.2.2. Design process of a capability model for learning analytics

IS are fundamental for LA (Nguyen et al., 2020; Rubel & Jones, 2016). For example, the digitalization of education is one of the drivers of LA (Ferguson, 2012). The RBV is used to explain how IS capabilities and Information Technology (IT) resources benefit organizations (Cosic et al., 2015). As we have the same goal, our study takes a resource-based perspective. The RBV is a prominent management theory and attributes the organizational performance to its ability to leverage resources (Barney, 1991; Grant, 1991). Capabilities refer to the organization's capacity to deploy resources to achieve the desired goal and are often developed by combining physical, human, and technological resources (Amit & Schoemaker, 1993). A key feature of capabilities is their purpose of enhancing the productivity of organizations' other resources (Makadok, 2001). Another key feature is that capabilities are embedded in organizations and cannot be easily transferred (Makadok, 2001). For HEIs willing to adopt LA, they can buy commodity assets like data warehouses and data science software but still need to develop their own LA capabilities. Resource-based capability models exist for various industries, e.g., manufacturing, retail, and health care (Adrian et al., 2018). However, no models grounded in the RBV exist in the LA domain yet.

We consider implementing LA to be a wicked problem. The interaction between subcomponents of the problem and its solution is complex, and the multidisciplinary nature calls for well-developed human social abilities. IS DSR principles (Hevner et al., 2004) are suitable for providing a solution to a wicked problem and are thus used to design a capability model for LA. Design science research projects comprise three cycles: the rigor cycle, the design cycle, and the relevance cycle (Hevner, 2007). The design of the capability model followed these cycles. In the rigor cycle, a systematic literature review was conducted to search the existing knowledge base for capabilities important for the implementation of analytics (Knobbout & van der Stappen, 2020a). Following the exaptation process (Gregor & Hevner, 2013), capabilities for successful business analytics and big data analytics were excapted to the LA domain. The systematic literature review started with 175 articles. After application of the inclusion and exclusion criteria, 15 key studies remained. The described capabilities and the ways they are operationalized were coded by two coders via open-coding principles (Strauss & Corbin, 1990). By synthesizing the results, a theoretical capability model was develop. Next, in the design cycle, the model was evaluated ex ante and refined (Knobbout et al., 2020). Empirical data was collected by conducting a single-case study at a Dutch institution that is mature in using analytics to improve education. During the study, interviews were held with two members of the analytics team as well as two end-users and archival data were collected. The interviews' verbalism transcripts were coded by two coders via axial-coding principles (Strauss & Corbin, 1990). Codes were derived from the theoretical model developed earlier. The case study led to several improvements to the model. Most noticeably, merging three capabilities and the renaming of three others. Also, improvements to most of the capability definitions were made. Now, in the relevance cycle, the final step is to ex-post evaluate the model's validity via field testing in the application domain.

### 7.2.3. Evaluation in Information Systems Design Science Research

Evaluation of the designed artifacts is a crucial step in DSR (Hevner et al., 2004; Peffers et al., 2012; Prat et al., 2015; Venable, 2010). A DSR process is a sequence of expert activities that produces a model and "[t]he utility, quality, and efficacy of a design artifact must be rigorously demonstrated via well-executed evaluation methods" (Hevner et al., 2004: p. 85). This means that the capability model must be studied and evaluated in the application domain (Hevner, 2007). To demonstrate the validity of a DSR project, four elements should be present: success of the corresponding artifacts, generality, novelty, and explanation capability (Carvalho, 2012). Hevner et al. (2004) provide guidelines that can be used as criteria and standards for the evaluation of design science research. However, there is consensus that not all these guidelines have to be followed to demonstrate the artifact's validity (Venable, 2010). What is important, is the need to address and solve a problem, have a clear design artifact, and perform some form of evaluation. Typical evaluation in IS research includes, among others, practice-based evaluation of effectiveness, usefulness, and ease of use (Prat et al., 2015). These evaluation types involve real practitioners as participants, the artifact's instantiation, and absolute measurement, i.e., the artifact is not compared to others. Common evaluation methods are expert evaluation, technical experiments, case studies, and illustrative scenarios (Peffers et al., 2012). To structure evaluation projects, Venable et al. (2016) provide a four-step approach: 1) explicate goals for the evaluation, 2) choose a strategy, 3) determine evaluation properties, and 4) design the evaluation episode. We used this approach to structure our evaluation study, as described in the next section.

## 7.3. Method

In this section, we describe our research goal and the research questions we will answer in our study. We then elaborate on the mixed-method approach we used. Next, we present the setup of our data collection, as well as the used research instruments. A pilot test was held to increase the research's rigor. Finally, we give an overview of the analyses that we performed on the collected data.

### 7.3.1. Research goal

As suggested by Prat et al. (2015), we applied commonly used styles of evaluation, namely practice-based evaluation of effectiveness and usefulness. First, our study's goal was to establish that the capability model works in a naturalistic situation, i.e., to measure its *effectiveness* (Venable et al., 2016). Second, since the model is used in the application domain, it is important to include its *usefulness* in the evaluation, i.e., the "degree to which the artifact positively impacts individuals' task performance" (Prat et al., 2015, p. 266). A third element for the evaluation of models relates to their *completeness* (Hevner et al., 2004; March & Smith, 1995). Completeness can be defined as the "degree to which the structure of the artifact contains all necessary elements and relationships between elements" (Prat et al., 2015, p. 266). Based on the aforementioned criteria, we constructed three research questions:

RQ1: Is the capability model effective, i.e., helpful to practitioners who want to implement LA at their institution?
RQ2: Is the capability model perceived useful by its users?
RQ3: Is the capability model complete?

To ensure the quality of our study, we proposed using different research methods. Mixed-method research increases the confidence in research data as weaknesses of one method are counterbalanced with another's strengths and vice versa (Thurmond, 2001). Different methods are applied to answer the three research questions: 1) pluralistic walk-throughs, 2) expert evaluation via a group discussion, and 3) a survey – see Table 26 and Figure 23.

| Research question | Method for Data Collection | Method for Data Analysis |
|---|---|---|
| RQ1: Is the capability model effective? | · Pluralistic walk-throughs | · Coding of roadmaps and transcripts |
| RQ2: Is the capability model perceived useful? | · Group discussions<br>· Survey (closed-ended questions) | · Coding of transcripts<br>· Descriptive statistics |
| RQ3: Is the capability model complete? | · Group discussions<br>· Survey (open-ended question) | · Coding of transcripts<br>· Coding of answers to open-ended question |

**Table 26:** Relation between Methods and Research Questions

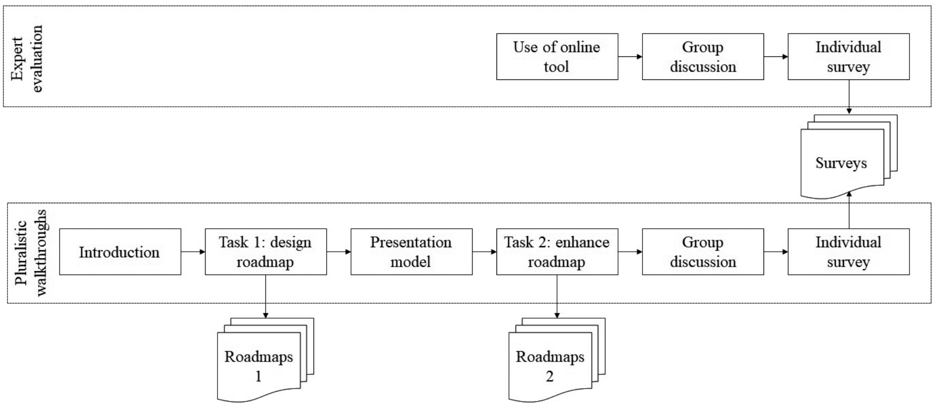


**Figure 23:** Research Methods and Data Collection.

### 7.3.2. Pluralistic walk-through

Human Risk & Effectiveness strategies are effective in situations where the evaluation's goal is user-oriented, it is possible to have real users in a real context, and time is limited (Venable et al., 2016). In an optimal situation, the implementation of LA with the capability model's support was studied over time in a real-world environment. However, in practice, such a process would take a long time. As time is constrained, we opted for a method that simulates the model's use in a real-world setting but faster. Prat et al. (2015) suggest to creatively and pragmatically generate new evaluation methods. In line with this suggestion, we used pluralistic walk-throughs. A pluralistic walk-through, also known as participatory design review (Kusters & Versendaal, 2013), allows for reviewing a product design by a group of stakeholders with varying competencies (Bias, 1994). This aligns with the model's goal, as stakeholder involvement is an important element in LA implementation (Hilliger et al., 2020). One of the benefits of a pluralistic walk-through is that it is rapid and generates immediate feedback from users, reducing the time necessary for the test-redesign-retest cycle (Thorvald et al., 2015). Pluralistic walk-throughs are often utilized to evaluate user interfaces and validate other sorts of IT artifacts (Emaus et al., 2010; Kusters & Versendaal, 2013). For instance, Dahlberg (2003) evaluates an application's design in its working context during five walk-throughs. Pluralistic walk-throughs are also conducted in the LA domain. For example, to support the design of a LA toolkit for teachers (Dyckhoff et al., 2012). In general, a pluralistic walk-through is defined by five characteristics (Bias, 1994):

1. Three types of participants are present: the designers of the system, usability experts, and users.
2. (Hard-copy) panels present the system.
3. All participants act like users.
4. Participants write down actions performed to complete given tasks.
5. After the walk-through, there is a group discussion.

Variations on the characteristics of pluralistic walk-throughs exist. Riihiaho (2002) separates users and experts' participation to give sufficient credit to the users' feedback and to save participants' time. We opted for the same approach and organized different meetings for novice practitioners (with no or limited experience with LA implementation) and experts (much experience with LA implementation). As usability is only one element we wish to evaluate, we decided to broaden the scope and invite LA experts rather than usability experts. Since the experts already know much about the capabilities necessary for LA and the capability model will provide them with fewer new insights than the novice practitioners, the novice practitioners' input can be collected separately. All sessions were hosted online via Microsoft Teams or Zoom to comply with COVID-19 regulations at that time.

### 7.3.3. Protocol for pluralistic walk-throughs

Based on the characteristics provided by Bias (1994), we conducted the following pluralistic walk-through for practitioners:

- We organized multiple sessions with stakeholders from different institutions as participants. In each session, the participants were from one, single organization. They were the capability model's target users – policymakers, senior managers, program directors, learning analysts, et cetera (characteristic 1). Each pluralistic walk-through lasted between two and three hours.
- To complete a planning task, the participants used the capability model. A digital version of the model was made available to support the task - see paragraph 3.8 (characteristic 2).
- During the pluralistic walk-through, participants were asked to solve a planning task: plan the implementation of a LA program at their institution to reach a predefined goal they have with LA (characteristic 3). This task led to a 'roadmap' in which the implementation process for the next two years was planned.
- During the pluralistic walk-throughs, at least one researcher represented the designers of the capability model. The researcher(s) took notes on interesting situations when the participants performed the task (characteristic 4). We also video-recorded each walk-through so they could be transcribed and analyzed.
- At the end of each pluralistic walk-through, there was a group discussion moderated by the researchers. During the discussion, participants elaborated on decisions made during the process and discussed whether the capability model provided sufficient support to complete the task (characteristic 5).

Similar to Rödle et al. (2019), we split the pluralistic walk-throughs in two phases. We did so to test whether the capability model gave participants new insights in what capabilities are important for the uptake of LA. The first phase was a baseline measurement to research what capabilities are already known and identified as important by the participants without the use of capability model. In the second phase, we introduced the model and measured what capabilities were present in the roadmap they made during phase one. To make the capabilities measurable, we formulated between one and three questions for each capability based on their definitions[13]. The participants were asked to analyze and enhance the roadmap they made in the previous phase using the capability model. This showed the effectiveness of the model (research question 1). Afterward, we discussed the outcomes as well as the model's usefulness and completeness with the participants (research questions 2 and 3). To steer the discussion, we presented three statements to the participants to react on:

- The capability model positively contributes to the adoption of LA by Dutch HEIs.
- The operational descriptions provided by the model help to make the adoption of LA more concrete.
- The capability model is complete. That is, there are no missing capabilities that are important to the adoption of LA by HEIs.

### 7.3.4. Expert evaluation

During expert evaluation, an artifact is assessed by one or multiple experts (Peffers et al., 2012). We invited experts that have multiple years of experience in implementing LA at their institution to participate in a group discussion. The experts were asked to use the capability model to gauge what capabilities are already present at their own institution and see which ones are still missing. Doing so gave them hands-on experience with the model. We then organized a group discussion to research whether they deem the capability model useful for practitioners who want to implement LA (research question 2) and whether they believe the model is complete (research question 3). During the session, a presentation on the development of the capability model was given first. Next, the outcomes of the experts' interaction with an online tool (see paragraph 7.3.6) were presented, followed by a group discussion. During this discussion, we asked the experts to react to the statements presented in paragraph 7.3.3.

### 7.3.5. Survey

Participants to both the pluralistic walk-throughs and the expert evaluation were asked to anonymously fill out an individual survey. Together with the data collected during the group discussions, the survey data is used to answer research question 2. In the survey, participants could score the capability model's usefulness, ease-of-use, and describe why it is (not) helping them implement LA at scale at their institution. To this end, we asked questions derived from the

---

13 Definitions can be found at found at https://hbo-kennisbank.nl/details/sharekit_hu:oai:surf-sharekit.nl:88cdbc33-c3d0-4748-b81d-263f6ad44876?q=LACM

Technology Acceptance Model (TAM)(F. D. Davis, 1985). TAM is used to measure peoples' intentions to use a specific technology or service and has been applied in the LA domain before (Ali et al., 2013; Rienties et al., 2018). In line with these studies, we focused on perceived usefulness and ease-of-use, ignoring external variables. Usefulness and ease-of-use proved to be important to the acceptance of technology in the educational domain (Herodotou et al., 2019) and capture the personal side of DSR (Prat et al., 2015). For our survey, we adopted the questions used in a previous study on the adoption of LA tools (Ali et al., 2013). Although we tested a model's usefulness instead of a LA tool's usefulness, we instantiated our model via a tool and deem the same questions relevant. Three questions were about the model's general perception, three questions about the perceived usefulness, and three about the ease-of-use. The survey used a Likert response scale of 1 (totally disagree) to 5 (totally agree). We added one open question on why the capability model is (not) helping participants implementing LA at their institution. As all participants in our sessions are Dutch-speaking, the original questions from Ali et al. (2013) were separately translated from English to Dutch by two researchers. Any discrepancies were discussed before finalizing the questionnaire. Surveys were distributed via Microsoft Forms. The survey questions are presented in appendix 3.



### 7.3.6. Participants

We now describe the participants' selection and characteristics for both the pluralistic walk-throughs and the expert evaluation.

#### *7.3.6.1. Participants for pluralistic walk-throughs*

A call for participation in our walk-through was made via SURF Communities[14], a Dutch website about innovation in the educational domain. The call was aimed at policymakers, IT staff, information managers, project managers, institutional researchers, and data analysts from HEIs with limited experience with LA but wanting to boost and scale up its use. We described the sessions we were planning and asked to get in touch with those interested in participating. Five institutions signed up to the pluralistic walk-throughs: one Dutch technical university, two Dutch universities of applied sciences, one Dutch institution for senior secondary vocational education, and one Belgian university of applied sciences.

14 https://communities.surf.nl/artikel/een-implementatieplan-voor-learning-analytics-opstel-len-met-het-lacm

#### *7.3.6.2. Participants for expert evaluation*

The experts all participate in a national program to boost the use of LA in the Netherlands. Via already-established contacts, we were able to organize a meeting with these experts. The project runs for two years now, but many experts have a long history of using LA. Therefore, they have profound knowledge and practical experience with the adoption of LA. During the session, seven experts from seven different institutions were present.

### 7.3.7. Analysis

When we determined what properties of the capability model to evaluate, we needed to consider in what way the collected data would be analyzed (Venable et al., 2016). With regards to the qualitative data, we analyzed what capabilities were present in the roadmaps made during the pluralistic walk-throughs. We did so via axial coding with codes derived from the capability model. Two researchers coded all statements made in the roadmaps and any discrepancies between coders were discussed until consensus was reached. To further enhance our research's quality, we verbatim transcribed the audio recordings of the pluralistic walk-throughs. These transcripts were also coded via axial coding and analyzed to see whether the capabilities used in the roadmaps reflect what was discussed during the roadmaps' design process. All coding was done in Atlas.ti. Coded data was used to research the frequency of appearance of the different capabilities. Moreover, we not only researched what capabilities are present in the roadmaps, we also investigated in what order they appear. Broos et al. (2020) distinguish four implementation phases. In the *initialisation phase*, LA goals and high-level planning are defined. Also, a project team is assembled. During the *prototyping phase*, instruments are developed and stakeholders are consulted. In the *piloting phase*, LA are deployed with real data of real users. The final phase is the *scaling phase*, in which LA are implemented at scale. Capabilities present in the roadmaps were classified accordingly to these phases.

The quantitative survey data was analyzed using descriptive statistics. There are two streams of thinking on whether Likert scales should be considered as ordinal data or interval data (Carifio & Perla, 2008). In line with other scholars within the LA domain (Ali et al., 2013; Rienties et al., 2018) and since our survey contains more than one Likert item, we deemed our data at interval level (Carifio & Perla, 2008). This allowed for calculating mean scores and standard deviations.

### 7.3.8. Pilot session

To increase a study's rigor, testing research instruments is an important step in IS research (Boudreau et al., 2001). Therefore, we performed one pilot session and used the experience gained during that session to enhance research materials, planning, tools, et cetera. Based on the pilot, some small adjustments were made, such as assigning a chairman/chairwoman and asking this person to lead the tasks at hand, formalizing discussion statements rather than asking open questions, or posting the assignments in the online chat so they could be read again. One major change was made to the way the roadmap is enhanced in the second phase of the pluralistic walk-through. For this end, we used an online tool. In this tool, the 818 operational descriptions from the capability model (see Appendix 2) are clustered in 138 different subgroups. Per subgroup, a question was formulated so the tool comprises 138 questions, and each one can be answered with either *Yes* or *No*. After answering, the tool calculates to what degree capabilities are already present and provides advice how to build the missing ones. However, during the pilot session, it became clear that the tool's questions were at a more fine-grained level than the roadmap itself. Also, due to the large number of questions, answering took a very long time. This made the walk-through tedious. Therefore, we decided to formulate questions at capability-level and use these during the next sessions. As we could not change the online tool on short notice, an Excel-file with drop-down menus, conditional formatting, and pivot tables was used instead to support the task of refining the roadmap. In the following pluralistic walk-throughs, this Excel-file was used.



## 7.4. Results

In this section, we present the results of our research. In total, 26 practitioners from five different educational institutions participated in the pluralistic walk-throughs. The modal number of participants is four per session. This is in line with other studies that use pluralistic walk-throughs as research method, where the number of participants lies between three and eighteen (see Dahl et al., 1995; Meixner et al., 2014). The characteristics of the participating institutions are described in Table 27. The institutions' names are substituted with fictional ones to ensure anonymity. Besides the pluralistic walk-throughs, we organized a group discussion with expert users to discuss the capability model. During this group discussion, seven experts were present and commented on the model. For our survey, we received 23 valid responses.

| Name | Type | Country | Intended improvement | Participants |
|---|---|---|---|---|
| Alpha | University of Applied Sciences | Belgium | Learning outcomes | 8 |
| Bravo | University of Applied Sciences | Netherlands | Learning environment | 4 |
| Charlie | University of Technology | Netherlands | Learning process | 4 |
| Delta | Institution for Senior Secondary Vocational Education | Netherlands | Learning process | 4 |
| Echo | University of Applied Sciences | Netherlands | Learning environment | 6 |

**Table 27:** Overview Participating Institutions

Please note that all sessions are held in Dutch and that all quotes are our translation. Participants to the pluralistic walkthroughs, experts, and respondents to the survey are named $P_x$, $E_x$, and $R_x$, respectively.

### 7.4.1. Effectiveness of the capability model

In each session, participants were asked to develop two roadmaps - the first one based on prior knowledge and own experiences, the second one enhanced with the help of the capability model. All five participating institutions made the first roadmap. However, it is important to note that not all institutions made the second, enhanced roadmap. The participants from Alpha decided that they wanted to discuss the new insights they got from the capability model before making a new roadmap. As one participant mentions, "*I think we would benefit more from it if we [make a new roadmap] more thoroughly on a later moment, as we have an important issue at hand with regard to our strategy*" (Alpha, P4). On the other hand, Bravo participants decided that the project they had envisioned should be run on a small scale first and large-scale implementation cannot be planned at this moment. The institution lacks a clear vision on the goals and application of LA and the participants deemed themselves incapable of making important decisions towards this end. They believe the organization should first decide on its strategy towards LA: "*What is the ethical framework, what is the overall goal? […] then you have a foundation you can further build on.*" (Bravo, P2). This indicates that institutions need to have at least the capabilities *Strategy, Ethics, Identifying benefits,* and *Policy & Code of Practice* developed to a certain degree before implementation at scale becomes possible. As described later on, the other pluralistic walk-throughs showed similar outcomes.

Aggregating the outcomes of all pluralistic walk-throughs, 88 capabilities were used to construct the roadmaps – see Figure 24. The roadmaps are axial coded based on the capabilities from the model. Initial coder agreement was 65%. In the walk-throughs' second phase, the participants got access to the capability model and supporting Excel-file to enhance their roadmap. This led to 28 more capabilities added to the second roadmaps. When we only consider the institutions that made a second roadmap - Charlie, Delta, and Echo - this is an increase of 45% (62 capabilities in their first roadmaps, 90 in their second).

| | | Alpha | Bravo | Charlie | | Delta | | Echo | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **Capability** | **1** | **1** | **1** | **2** | **1** | **2** | **1** | **2** | **Total** |
| Data | Data usage | | 1 | 3 | 3 | 2 | 2 | | | 11 |
| | Feedback on analytics | | | | | | | | | 0 |
| | Quality | | | | **1** | | | | | 1 |
| | Reporting | | | | | 2 | 2 | | | 4 |
| | Sourcing & integration | | 3 | 3 | 4 | 4 | 7 | | | 21 |
| Management | Capability development | 2 | 1 | | **2** | | | | 1 | 6 |
| | Culture & readiness | 1 | 1 | | | | | 1 | 1 | 4 |
| | Evidence-based & theory-driven | | | | | 1 | 1 | 1 | 1 | 4 |
| | External environment | | | | | | | 1 | 1 | 2 |
| | Funding & investment | | | 1 | 1 | | | | | 2 |
| | Identifying benefits | 1 | | 1 | 1 | | | | | 3 |
| | Implementation Deployment & Application | 3 | 5 | 6 | 6 | 7 | 7 | 6 | 7 | 47 |
| | Performance monitoring | | 1 | | | 1 | 1 | 1 | 1 | 5 |
| | Policies & CoP | 1 | | 1 | 2 | | | 1 | 1 | 6 |
| | Responsibility & accountability | | | 1 | 1 | | | | | 2 |
| | Strategy | 1 | | | | | | | **1** | 2 |
| People | Collaboration | | | | | | | | | 0 |
| | Combined skills and knowledge | 1 | | | | | | | | 1 |
| | Communication | | | 1 | 2 | | | | | 3 |
| | Stakeholder Identification & Engagement | 1 | 1 | 3 | 5 | | 7 | 2 | 2 | 21 |
| | Training | | | 2 | 2 | | | | | 4 |
| Privacy & Ethics | Ethics | | | | **1** | | **1** | | **1** | 3 |
| | Human decision-making | | | | | | | | | 0 |
| | Legal compliance | | | 1 | 1 | 2 | 3 | | | 7 |
| | Security | | | | | | | | | 0 |
| | Transparency | | | | | | | | | 0 |
| Technology | Automation | | | 1 | 1 | | | | | 2 |
| | Connectivity | | | | | | | | | 0 |
| | Infrastructure | 1 | 1 | 2 | 5 | 2 | 2 | 2 | 2 | 17 |
| | System characteristics | | | | | | | | | 0 |
| | Total | 12 | 14 | 26 | 38 | 21 | 33 | 15 | 19 | 178 |

**Figure 24:** Codes assigned to the First Roadmaps (1) and the Second, Enhanced Roadmaps (2). Capabilities Only Present in Second Roadmap are Shaded and Printed in Bold.

The capability *Implementation, deployment & application* is the most frequently present. This is logical, as this capability relates to the planning of LA implementation at an institution, which is exactly the task during the pluralistic walk-throughs. Further analysis shows that not all capability categories are present in all roadmaps. For example, the category *Data* is absent in the roadmaps of Alpha and Echo. Moreover, the category *People* is (almost) missing from the roadmaps of Bravo, Delta, and Echo. Most noticeably, however, is the absence of the category *Privacy & Ethics*. This category was completely absent in the roadmaps of Bravo, Delta, and Echo, and was only mentioned once in the roadmap of Alpha.

Our study aims to investigate whether the capability model helps stakeholders identify capabilities that are overlooked in an implementation roadmap. Therefore, we analyze what *unique* capabilities can be distinguished, i.e., capabilities entirely absent in the first roadmaps but added to the enhanced ones. These are presented as bold, shaded numbers in Figure 24. All institutions added *Ethics* to their roadmaps. When asked why no capabilities from the category *Privacy & Ethics* were initially planned, participants say that this was already present or discussed within the organization: "*The ethical part is described in our vision on data. […] For learning analytics, there is a privacy statement*" (Echo, P5), "*the discussion about ethics is an old one*" (Delta, P3), or "*we have been busy to get [privacy and ethics] organized but I think it is not mentioned in our discussion*" (Charlie, P4). This endorses the observation of Bravo that – at least some – ethical considerations must be decided on before starting the implementation. Nonetheless, the participants agreed this topic should have a prominent place in the implementation process. As a result, capabilities regarding ethics were added to the second, enhanced roadmaps of all institutions. Moreover, two institutions added capabilities regarding *Capability management* to their roadmap, while *Quality* and *Stakeholder identification & engagement* were each included in one roadmap. Although this does not sound much in absolute numbers, without these capabilities, successful adoption of LA is placed in serious jeopardy. For example, involving stakeholders is crucial for effective implementation (Hilliger et al., 2020). The capability model helps identify capabilities that are missing in the roadmap: "*you immediately see which topics got a lot of our attention and which one got less*" (Echo, P1).While analyzing the roadmaps, it became apparent that many capabilities were mentioned during the sessions but not all of them were added to the roadmaps themselves. Axial coding of the transcripts supports this observation. As shown in Figure 25, many capabilities were discussed but not made explicit in the roadmap. Remarkably, capabilities from the category *Data* are discussed by Alpha and Echo but none of the capabilities from this category are adopted in the roadmap. This shows that the capability model serves two goals: 1) inform practitioners about capabilities that are yet unknown to them, and 2) support users making explicit choices about what capabilities to adopt in their plan towards LA implementation. The latter is also expressed by one of the participants: "*for some cases [the need for certain capabilities] might be implicit but it would be good to make them explicit, so they can be thought about. And if you think that they are not applicable, you need to explain why you think so.*" (Delta, P3).

| | | Alpha | | Bravo | | Charlie | | | | Delta | | | | Echo | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Capability** | T | R | T | R | T1 | R1 | T2 | R2 | T1 | R1 | T2 | R2 | T1 | R1 | T2 | R2 | **Total** |
| Data | Data usage | 6 | | 6 | 1 | | 3 | 2 | | | 2 | 2 | | 5 | | 1 | | 28 |
| | Feedback on analytics | 1 | | | | | | 1 | | | | | | | | 1 | | 3 |
| | Quality | 4 | | 3 | | | | | 1 | | | | | 2 | | 1 | | 11 |
| | Reporting | 5 | | 3 | | 3 | | | | 2 | 2 | | | | | | | 15 |
| | Sourcing & integration | 8 | | 13 | 3 | 2 | 3 | 3 | 1 | 5 | 4 | | 3 | 7 | | 3 | | 55 |
| Management | Capability development | | 2 | 1 | 1 | | | | 2 | | | | | | | | 1 | 7 |
| | Culture & readiness | | 1 | 4 | 1 | 2 | | 5 | | 5 | | 3 | | 1 | 1 | | | 23 |
| | Evidence-based & theory-driven | | | 1 | | 1 | | | | 3 | 1 | 1 | | 1 | 1 | | | 9 |
| | External environment | | | | | 1 | | | | 4 | | 4 | | | 1 | | | 10 |
| | Funding & investment | 1 | | 2 | | 2 | 1 | 1 | | 2 | | 3 | | 1 | | | | 13 |
| | Identifying benefits | 2 | 1 | 3 | | 3 | 1 | | | 4 | | | | 5 | | | | 19 |
| | Implementation Deployment & Application | 11 | 3 | 14 | 5 | 3 | 6 | 2 | | 11 | 7 | 2 | | 16 | 6 | 5 | 1 | 92 |
| | Performance monitoring | | | 1 | 1 | 1 | | | | 4 | 1 | 1 | | 1 | 1 | 1 | | 12 |
| | Policies & CoP | 1 | 1 | | | | 1 | 2 | 1 | 1 | | 1 | | 2 | 1 | | | 11 |
| | Responsibility & accountability | 2 | | 2 | | | 1 | 1 | | | | 3 | | 4 | | 3 | | 16 |
| | Strategy | 1 | 1 | | | 1 | | | | 2 | | | | 4 | | 1 | 1 | 11 |
| People | Collaboration | 3 | | 4 | | 2 | | 1 | | 4 | | 1 | | 4 | | 2 | | 21 |
| | Combined skills and knowledge | | 1 | | | 1 | | | | 1 | | | | 1 | | | | 4 |
| | Communication | 2 | | 2 | | | 1 | | 1 | 1 | | | | | | 1 | | 8 |
| | Stakeholder Identification & Engagement | 6 | 1 | 7 | 1 | | 3 | 3 | 2 | 7 | | 2 | 7 | 2 | 2 | 4 | | 47 |
| | Training | | | | | | 2 | | | 3 | | | | | | | | 5 |
| Privacy & Ethics | Ethics | | | 2 | | | | 2 | 1 | | | 3 | 1 | 1 | | 1 | 1 | 12 |
| | Human decision-making | | | | | | | | | | | | | | | | | 0 |
| | Legal compliance | 1 | | 6 | | 3 | 1 | 3 | | | 2 | 1 | 1 | | | 1 | | 19 |
| | Security | | | | | | | | | | | | | 1 | | 1 | | 2 |
| | Transparency | | | 1 | | | | | | | | | | | | | | 1 |
| Technology | Automation | | | 2 | | | 1 | | | | | | | | | | | 3 |
| | Connectivity | | | | | | | | | | | | | | | | | 0 |
| | Infrastructure | 4 | 1 | 5 | 1 | | 2 | 1 | 3 | 6 | 2 | 2 | | 5 | 2 | 2 | | 36 |
| | System characteristics | | | 2 | | | | 1 | | | | | | | | | | 3 |
| | Total | 58 | 12 | 84 | 14 | 25 | 26 | 28 | 12 | 65 | 21 | 29 | 12 | 63 | 15 | 28 | 4 | 496 |

**Figure 25:** Appearance of Codes in Transcripts (T) and Roadmaps (R).

Additional to researching the capabilities' presence in the roadmaps, we can investigate in what order they appear. Participants of the pluralistic walk-throughs are asked to plan LA implementation at their institution and to consider the timeliness and sequence of capability development. Based on this sequence, we can research what capabilities play a role in which implementation period. Similar to Broos et al. (2020), we distinguish four phases (initiation, prototyping, piloting, scaling). Using the definitions provided by Broos et al, we mapped the capabilities mentioned in the roadmaps accordingly. See Figure 26 for the capabilities' classification per phase. Our data shows that certain capabilities are important to multiple phases, e.g., *Implementation, Deployment & Application* and *Legal Compliance*. These capabilities play a role throughout the planning, design and

7

deployment of LA. Others are particularly important during the initialization of LA, e.g., *Ethics* and *Strategy*. These capabilities must be developed at the very start as they define the scope of LA' use in the institution. When real end-users come into play during the pilot phase, capabilities like *Communication* and *Reporting* must be present. At that point, the analyses' outcomes are presented to an actual audience, thus the need for capabilities that relate to these elements. Finally, there are also capabilities that can be developed during the scaling phase, e.g., *Automation* and *Training*. They have limited use when LA is applied at a small scale but become necessary when scaling up on its use. Although this classification is conceptual and needs further development, its shows there is a certain order in which capabilities must be developed to effectively adopt LA.

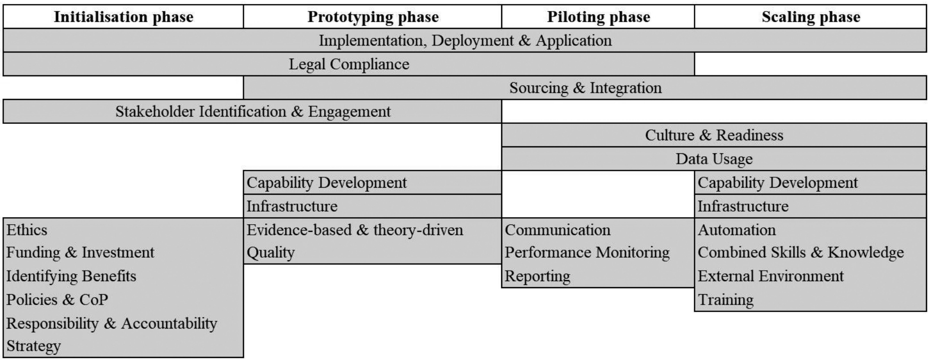


**Figure 26:** Sequential Classification of Capabilities from Roadmaps

### 7.4.2. Perceived Usefulness of the capability model

To research the capability model's perceived usefulness, we held group discussions with both the participants to the pluralistic walk-throughs and the expert evaluation. Participants are asked to react to statements regarding the model's positive impact and the operational descriptions it provides (the first and second statement presented in paragraph 7.3.3).

The experts reacted positively to the statement that the capability model supports the adoption of LA at Dutch HEIs. They believe that the model provides an overview of important aspects. However, the detailedness of the model might serve the self-fulfilling prophecy that when institutions think LA is too complex to implement, the model supports this feeling. This became apparent during the pluralistic walk-throughs, when no enhanced roadmap was made by participants from Bravo for they believed certain capabilities, e.g., *Strategy* and *Policy & Code* *of Practice,* must developed first. According to the experts, the capability model's added value lies in the discussion practitioners have when using the model together to plan LA implementation. As one expert mentions, "*the value lies in the dialog*" (E2). Furthermore, the experts think that the operational descriptions are supportive. Sometimes, a capability is already partly developed, but operational descriptions help identify weak spots and enable institutions to develop this particular capability further. The participants to the pluralistic walk-throughs were also positive about the model. According to them, it helps to get insights into the aspects important for LA implementation: "*there are some things you don't think about spontaneously*" (Alpha, P4) and "*[t]he model helped me to see clearer [whether we] thought well enough about all main points*" (Charlie, P2). The model "*contributes to a better plan and a more complete plan*" (Delta, P3). This reflects the observation made earlier that, although many capabilities were discussed, only a small portion was included in the roadmaps. The model can therefore help to make a more explicit plan.

Besides group interviews, we used a survey based on the Technology Acceptance Model to research the model's perceived usefulness and ease-of-use. The survey contains three groups of closed-ended questions: questions 1 to 3 measure *overall perception*, questions 4 to 6 measure *perceived usefulness*, and questions 7 to 9 measure *ease-of-use*. The tenth question is an open-ended question regarding the effectiveness of the capability model. We report the mean (M) and standard deviation (SD) in the text for each factor. The outcomes per question are shown in Table 28. In line with Rienties et al. (2018), we take 3.5 or higher as positive cut-off value and anything lower than 3 as negative cut-off value. We regard scores between those numbers as neutral. In total, we received 23 valid responses (n = 23).

The *overall perception* of the model is positive (M = 3.8, SD = 0.3). The questions whether respondents think the capability model is useful and whether they would like to use it in their work are rated highly positive. The question whether the model is more useful than comparable models scores neutral. However, this is probably due to the limited experience with other models, letting respondents answer neutral to this question.

The c*apability model's perceived usefulness* is positive (M = 4.1, SD = 0.5). The respondents think the model enables them to get insight into the capabilities necessary for the successful implementation of LA and that the model helps to identify what capabilities need to be (further) developed at their institution.

Although still positive, the third item in this factor ranks the lowest: the respondents score the question whether the capability model provides them with relevant information in what way to operationalize LA capabilities with a 3.7 (SD = 0.8). This might result from the large number of operational descriptions (818) in the model and the relatively short duration of the sessions (between two and three hours). Due to this combination, participants were probably not able to fully use the model's fine-grained constructs.

The third and last measured factor, *ease-of-use*, scores the lowest, yet positive (M = 3.5, SD = 0.6). Although respondents think the capability model is easy to understand and its use is intuitive, 35% of respondents believe it overburdened with information. This is also expressed in answers to the open question: "*the model [is] extensive and therefore not clear*" (R5) and "*[a]t first sight, there are a lot of questions and there is a lot of information*" (R8).

| Question | Mean | Std. Dev. | Min. | Max. |
|---|---|---|---|---|
| Q1: Useful | 4,2 | 0,4 | 4 | 5 |
| Q2: Use in work | 4 | 0,4 | 3 | 5 |
| Q3: Comparison other models | 3,2 | 0,4 | 3 | 4 |
| Q4: Insights | 4,3 | 0,5 | 3 | 5 |
| Q5: Identify | 4,3 | 0,5 | 4 | 5 |
| Q6: Operationalize | 3,7 | 0,8 | 2 | 5 |
| Q7: Easy to understand | 3,9 | 0,8 | 2 | 5 |
| Q8: Intuitive | 3,6 | 0,7 | 2 | 5 |
| Q9: Overburden with information | 2,9 | 0,9 | 1 | 4 |

**Table 28:** Descriptive Statistics per Survey Question

In the answers to the open questions, most respondents write that they think the capability model is beneficial and useful. The model, for example, "*provides a clear framework, insights in what is necessary, as well as what steps to take*" (R20) and it "*offers mind-sets to develop policy further*" (R15). Similar to what the experts mentioned in the group discussion, some people think that the model can serve as a checklist (R12) and believe it helps to start conversations with other stakeholders (R23).

### 7.4.3. Completeness of the capability model

To answer the third research question, we researched the capability model's completeness, i.e., whether all capabilities important to LA adoption are present.

Data collected during the group discussions as well as via the open-ended survey question were used to this end.

Asked to comment on the model's completeness (the third statement presented in paragraph 3.3), the experts think that the pedagogical interventions that follow on the analysis and visualization of learner data are yet underexposed. The model comprises a capability to monitor the performance of the analytical processes and applications (*Performance monitoring*) but this does not include the measurement of improvements made to education. The experts suggest adding this to the model. The participants to the pluralistic walk-throughs also mention some potential shortcomings. Most prominent, the position of pedagogical use of LA in the capability model is not apparent enough. That is, the role of remediation is underexposed (Alpha, P1) and it is unclear how to take into account the contextual setting of education (Charlie, P1). Although these aspects are part of the capabilities *Implementation, Deployment & Application* and *Evidence-based & Theory-driven*, respectively, this might not be clear enough.

Other improvements to the model are suggested in the answers to the open question in the survey. Two respondents think the capability model alone is not enough to put LA into practice. As one respondent puts it: "*additional materials can be developed that focus on the HOW. The current model mainly visualizes the WHAT*" (R8). This calls for the development of additional materials like templates with concrete advice on how to plan the capabilities' development. Another respondent suggests adding elements specific to education to the model, as the current model is "*quite generic and could also be applied to other analytical domains*" (R18). This is in line with comments made in the group discussions with both practitioners and experts. That the model is generic is - at least to a certain degree - to be expected as the first version of the model is exapted from literature from the big data analytics and business analytics domains (Knobbout & van der Stappen, 2020a). Nonetheless, it highlights the need to better position the role of educational context and theory in the model.



## 7.5. Discussion

The research's main goal was to evaluate the capability model in its application domain. Our study meets the four elements that demonstrate the validity of DSR (Carvalho, 2012): successfulness, generality, novelty, and explanation capability. The at-scale adoption of LA at HEIs is a wicked problem. The model is designed to support practitioners such as program managers, policymakers, and senior management by providing 1) an overview of necessary capabilities and 2) insight into how to operationalize these capabilities. This study shows that the capability model is *successful* in solving the wicked problem of LA adoption. Although the model is designed for Dutch HEIs in particular, its *generality* is proved by including a Belgian HEI and an institution for senior secondary vocational education in the evaluation process. The frequency and types of capabilities mentioned in these institutions' roadmaps do not differ from the participating Dutch HEIs. The successful pluralistic walk-throughs with participants from these two divergent institutions shows that the model is also useful in other contexts. Moreover, the capability model is *novel*, as it is the first model grounded in the RBV and it is innovative in its approach to tackle the challenges faced by HEIs that wish to implement LA. Finally, the model is *explanatory* because it describes what resource-based capabilities are necessary for the adoption of LA, how to operationalize these capabilities, and what capabilities must be developed in what phase of the implementation process. This knowledge was absent prior to our research.

The remainder of this section discusses the outcomes of our work, its relevance for research and practice, the study's limitations, and provides recommendations for future research.

### 7.5.1. Research outcomes

The study's first research question investigates the capability model's effectiveness to practitioners who are in the process of implementing LA at their institution. Our research participants represented different stakeholders at educational institutions, which is important for successful LA adoption (Hilliger et al., 2020). The model proved effective in the planning of LA implementation. Our analysis showed that many capabilities are often overlooked when creating implementation roadmaps. Most noticeably, there was almost no attention for privacy and ethics. This is remarkable as these are important aspects of LA (Gašević et al., 2016; Pardo & Siemens, 2014). Using the model, participants were better able to make a more complete and more explicit roadmap. In the second phase of the pluralistic walk-throughs, the capabilities' inclusion to the enhanced roadmaps increased with 45%.

The second research question regards the capability model's perceived usefulness. During the group discussions, both expert and practitioners were positive about the model's scope and depth. The survey outcomes support this conclusion: both the overall perception and perceived usefulness score well. However, due to its large size, the ease-of-use of the model is mediocre. Especially when we used the online tool with 138 different questions in our first, pilot pluralistic walk-through, the participants might have been overwhelmed. This has to do with the *instantiation validity* of the model, i.e., “the extent to which an artifact is a valid instantiation of a theoretical construct or a manifestation of a design principle” (Lukyanenko & Parsons, 2015, p. 430). Due to the complexity of the model’s instantiation, the model itself is perceived hard to use. We plan to enhance this during future research (see paragraph 5.3).

The third research question considers the capability model’s completeness. All the important capabilities are present in the model. Nonetheless, some improvements can be made. First, the scope of *Performance monitoring* must be broadened. Its current definition sounds, “*In what way the performance of analytical processes and applications are measured.*” This definition is too narrow. The pluralistic walk-throughs’ outcomes show that the performance of processes and systems must be monitored and the interventions’ performance should be measured. This is in line with opinions expressed by other scholars (Baker, 2019; Foster & Francis, 2019; Knight et al., 2020). Second, the position of pedagogical theory and context and the evaluation of the impact of interventions should be made more prominent. This is an essential conclusion as the benefits of LA to teaching and education are often overlooked (Hilliger et al., 2020). Lastly, some participants suggest the development of supporting materials to enhance the practical use of the model.



### 7.5.2. Implications for research and practice

In this study, we showed that the capability model is comprehensive and helps to identify capabilities overlooked in the implementation planning, make implementation planning more concrete, that the model is perceived useful by practitioners, and that the model is complete. HEIs face many challenges when implementing learning analytics (Tsai & Gašević, 2017b). By using, our model overcomes these challenges and helps the LA domain to enhance educational practice. In addition, other domains might also benefit from this research. One pluralistic walk-through participant has many experience with data analytics in the energy industry and the financial industry. He mentions that “*it is all relevant when assembling any data analytics team*” (Echo, P2). This comment shows that our model’s relevance goes beyond the LA field and might also be applicable in other domains.

This study's main contribution to the IS domain is the evidence that pluralistic walk-throughs can be conducted to ex-post evaluate an IT artifact. There is debate about what method is suitable for the evaluation of what type of artifact (Venable et al., 2016; Winter, 2008). However, no hard rules exists and some scholars argue that the evaluation must be designed pragmatically (Prat et al., 2015), encouraging researchers to generate creative, new methods. In our research, we needed a method that simulates the capability model's use in practice by real users but that could be applied in a limited amount of time. Pluralistic walk-throughs are often applied for ex ante evaluation when designers are still willing and able to make changes to the design (Riihiaho, 2002). By adapting this method, we were able to use it for ex-post evaluation as well.

### 7.5.3. Limitations

Our research has several limitations. First, since participants voluntarily joined our sessions, there might be non-response or self-selecting bias. This might affect the research outcomes as it is impossible to determine what differences exist between our sample and the population (Bryman & Bell, 2015). However, we mitigated this effect by having a well-balanced mix of participating institutions and people: 26 practitioners who represented five different educational institutions that are comparable with other institutions in the Netherlands. This poses another limitation. As we intentionally focused on Dutch-speaking participants and thus institutions, outcomes might not be generalizable to institutions in different countries. Further research is necessary to address this shortcoming.

The quality of survey items is another caveat of our research. In survey research, the validity of a questionnaire is analyzed via factor analysis. However, this requires a large enough sample size. A common rule is to have 10 to 15 respondents per variable (Field, 2005), which in our case (9 questions) means between 90 and 135 respondents. We only have 23 respondents, thus validity could not be statistically established. Moreover, the question whether the capability model is more useful than comparable models might score lower than expected due to participants' limited experience with other models. Two respondents explicitly state that they do not know other models and therefore cannot answer the question. The current questionnaire, however, requires an answer. When we remove this question from the analysis, *overall perception* of the model raises to 4.1 (SD = 0.3). Although the survey was tested before use, this issue was overlooked but should be considered in future research.

### 7.5.4. Future work

In this research, we evaluated a capability model for LA. During empirical validation, the model proved effective and useful to practitioners. Nonetheless, some improvements can be made to the presentation of the model. Most importantly, the position of pedagogical theory and the measurement of the interventions' effect on learning can be enhanced. Furthermore, in the survey, the model scores relatively low on the factor *ease-of-use*. Some of the participants request additional materials such as templates or tools to support the capability model's use. Future work could focus on the development of supporting material, which is also done for e.g., the SHEILA framework (Tsai et al., 2018). The current model helps users identifying important capabilities, but it lacks steps how to develop these capabilities. Considering the different types of models (Broos et al., 2020; Colvin et al., 2017; Dawson et al., 2018), the capability model is an input model and not a process model. Nonetheless, by analyzing the order in which the capabilities appeared in the roadmaps, we can distinguish certain patterns. Some capabilities are important at the start of the implementation process while others become necessary later on. Knowing what capabilities must developed in the near future helps allocating resources to achieve this development. For example, funding might be secured well in advance to hire experts to design functional LA systems. Although our results with respect to this topic are yet conceptual, it provides some initial insights in the order capabilities must be developed in. Further research towards this direction can help transforming the current input model into a process model. This is beneficial, as process models are more suitable for capturing the complexity of higher education than input models (Broos et al., 2020). A last direction for future work is the capability model's generalizability to educational systems outside the Netherlands. We invite other researchers to use the model in other settings and report on the outcomes.

# CHAPTER

# 8

# 8. General Discussion

Learning analytics aim at improving learning and the learning environment via the analysis of student data (Duval, 2012; Ifenthaler, 2015; LAK, 2011). To support learning analytics at higher educational institutions, scholars have developed several implementation models (Arnold, Lynch, et al., 2014; Bichsel, 2012; Broos et al., 2020; Ferguson et al., 2014; Greller & Drachsler, 2012; Norris & Baer, 2013; Siemens et al., 2013; Tsai et al., 2018). However, these extant models have their limitations. Most notable, the way how to operationalize them at scale remains unclear, their empirical validity is often not proved, and they are mainly focused on the educational context in North America, the United Kingdom, and Australia. To overcome these shortcomings, we developed the Learning Analytics Capability Model. The model is the answer to the main research question stated in the introduction of this Ph.D. thesis:

*"Which Information System-related capabilities for learning analytics benefit teachers and learners in Dutch higher educational institutions?"*

In this final chapter, we first summarize the outcomes of the studies that were conducted to justify, design, and evaluate the Learning Analytics Capability Model. Second, we discuss the studies' outcomes in relation to the main research question. Next, we describe our work's contributions and implications to both academics and practitioners. Lastly, we describe the Ph.D. research's limitations and provide directions for future research.

## 8.1. Main findings

In the Ph.D. thesis's first part, we described what learning analytics is and posted the research question:

*"In what way does existing literature on learning analytics interventions operationalize affected learning?"*

During a systematic literature review, seven databases were searched for learning analytics literature that describes in what way to assess the effects of learning analytics on learning. From 1932 search hits, 62 key studies were identified as relevant. Operational definitions of affected learning were classified into three categories: *Learning environment, Learning process*, and *Learning outcome*. Most key studies present results that relate to the latter: containing 42 key studies, *Learning outcome* is by far the largest category. A deepening analysis yielded a refined classification scheme with eleven subcategories – see Figure 27. Data such as grades, test scores, and virtual learning environment log files are often available for researchers at higher educational institutions. This is reflected by the fact that most key studies related to the subcategories *Learning process – Online activity & behavior* and *Learning outcome – Knowledge and skills*. However, since grades are only a proxy for learning (Teaching Commons, 2019), it would be beneficial to include other learning measures. Although now only found in nine key studies, in line with Joksimović et al. (2018) we argue that cross-categorical measures lead to a more complete view on learning.

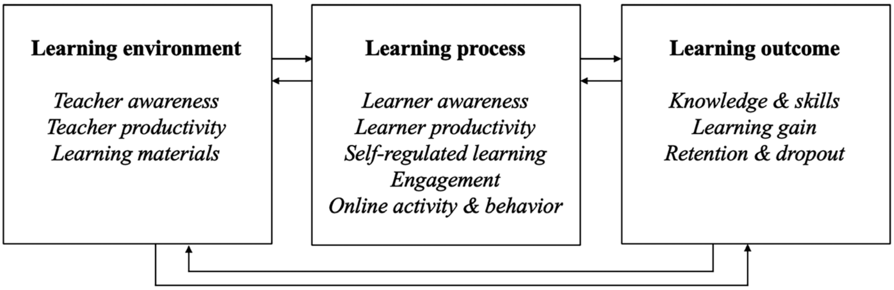


**Figure 27:** Refined classification scheme for operational definitions of learning affected by learning analytics interventions



The Ph.D. thesis's second part justifies the research's need and relevance. It starts with the research question:

*"What issues are encountered when implementing an experimental learning analytics tool in the case organization's virtual learning environment?"*

The research question led to a classification of the case organization's issues when implementing a learning analytics tool. The classification was done via the learning analytics framework by Greller and Drachsler (2012), who distinguish six critical dimensions: *Data, Objectives, Stakeholders, Internal limitations, External limitations*, and *Instruments*. The dimensions are critical in the sense that learning analytics designs need instantiations from each dimension. The tool's implementation was successful in the sense that it allowed visualizing data that was previously inaccessible for students and teachers and provides valuable insights into what is necessary for learning analytics adoption at the institution. However, there was room for improvement. The case study's results showed that five of the six critical dimensions encountered problems. For example, involved teachers mentioned that they tracked too few activities. As a result, pedagogical interventions could not be designed effectively. Also, per student only learning activities performed on one device were tracked due to a technical error. Another problem related to the self-selection of better-performing students. Finally, the connection between the organization's systems and the tool could only be established at the last moment as stakeholders from the IT department were involved too late. Future research was planned but this could not be conducted since the tool was terminated due to privacy issues. Although this became a problem when the initial experiment was already finished, it shows that privacy is an important element for learning analytics. The case study showed that implementing learning analytics – even at a small scale – is a challenging endeavor and many organizational capabilities need careful consideration.

A second case study provided additional support to the Ph.D. research. It focuses on the importance of clean data and answered the following research question:

> *"What are the effects of (unspoken) choices made during the cleaning process of student data on the outcomes when these data are in turn used for learning analytics?"*

The study's original idea was to answer three learner-related questions based on data collected via Moodle during six blended courses. However, the dataset had many missing values and choices had to be made to clean the data. Extant literature often does not describe the need for data cleaning and what choices are made while doing so, even though cleaning and munging data is an important part of conducting data science (VanderPlas, 2016). Of 438 analyzed papers from the Learning Analytics & Knowledge (LAK) conference proceedings (2011 till 2018), only 17 papers describe either the cleaning or preprocessing of learner data. In our study, the raw data were used to construct 12 datasets, each of them based on realistic choices and assumptions. Analyzing students' interaction with the virtual learning environment using the different datasets led to different outcomes. For example, based on one dataset, a student spent 45% of his/her study time on course A. Using another dataset, the same student spent 28% of study time on this course. Since learning analytics are used to intervene on the analyses' outcomes, results not reflecting reality might lead to wrong interventions or distrust in learning analytics in general. Therefore, having mechanisms in place to correctly clean data and be transparent about choices made during the process are important capabilities for higher educational institutions that want to implement and use learning analytics.

The thesis's third part described the design of a resource-based capability model for learning analytics – the Learning Analytics Capability Model. By conducting a systematic literature review, the following research question was answered:

*"What capabilities for the successful adoption of learning analytics can be identified in existing literature on big data analytics, business analytics, and learning analytics?"*

As the large-scale learning analytics implementation is a nascent topic in existing learning analytics literature, we included literature from more mature research fields such as business analytics and big data analytics. This way, capabilities important for analytics in those domains are *exapted* to the learning analytics domain (Gregor & Hevner, 2013). In total, 15 key studies were analyzed. 461 ways to operationalize analytical capabilities were extracted from the found literature. The operational definitions from business analytics and big data analytics literature were first grouped based on open coding principles to construct the model. This led to 23 capabilities divided over four second-order capabilities (categories): Data, Management, People, and Technology. Next, the operational definitions from learning analytics literature were coded based on axial coding principles. Codes corresponded with the capabilities found earlier in the research. Not all operational definitions from the learning analytics literature fit the 23 capabilities, so 11 more capabilities were formulated via open coding. Six of these new capabilities fitted the existing categories. On the other hand, five capabilities were placed in a new category: Privacy & Ethics. Surprisingly, this category is largely unmentioned in business analytics and big data analytics literature, since personal and privacy-sensitive data is also handled in those domains. Although data privacy and individual rights are important factors in big data and business analytics ecosystems (Pappas et al., 2018), this is not reflected in existing capability models for these systems. However, related capabilities are present in learning analytics models and thus included in the model. As a result, the first, theoretical Learning Analytics Capability Model comprises 34 capabilities grouped into five categories.

The next research question aims at evaluation and refining the theoretical model:

*"How can the Learning Analytics Capability Model be evaluated and refined based on empirical data from a single Dutch higher educational institution that is mature in the use of learner data to improve learning?"*

Both interview data and archival data were collected during a case study at a Dutch university that uses learner data to improve learning and learning environments in particular. Four interviews were held: two with members of the analytics team (project leader and data engineer) and two with end-users (policymaker and academic advisor). During the interviews, the capabilities from the first, theoretical Learning Analytics Capability Model were discussed. The interviews were verbatim transcribed and coded along with the model's capabilities. The analysis showed that capabilities excapted from the business analytics and big data analytics domains are indeed relevant for higher educational institutions that implement learning analytics. Capabilities like *Sourcing & Integration*, *Automation*, and *Training* were absent in extant learning analytics literature but present at the case organization. Nonetheless, several refinements could be made to the model. First and foremost, four capabilities were merged with other capabilities already present in the model. Merging was done as some capabilities can be regarded as useful when combined. For example, skills can only be developed with the right knowledge. The theoretical model contained three related capabilities: *Skills, Knowledge*, and *Combined Skills & Knowledge*. Based on mutual dependability, these three were merged into one. For clarity reasons, some capabilities were renamed. For instance, *Market* is a term adopted from business analytics, which is a more commercial-oriented field. It relates to the environment outside an organization, including customers, competitors, and government. However, there is often collaboration with other higher educational institutions or other non-profit organizations in the learning analytics domain. Therefore, the capability *Market* was renamed *External Environment* so it better suits the educational environment. Finally, changes to the capabilities' definitions were made based on the collected data – see Appendix 1. The resulting Learning Analytics Capability Model now comprises 30 capabilities grouped into five categories – see Figure 28.

The thesis's fourth and last part elaborated on the Learning Analytics Capability Model's ex-post evaluation in the application domain. It provides an answer to the research question:

*"How to ex-post evaluate the Learning Analytics Capability Model in the context of Dutch-speaking educational institutions?"*

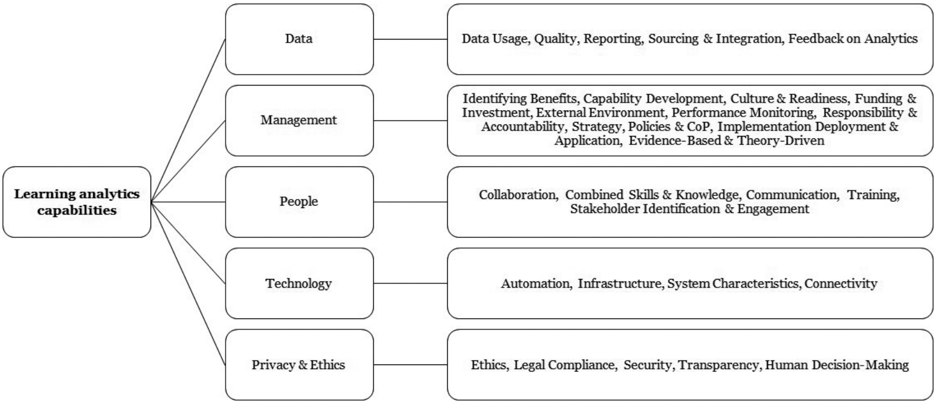


**Figure 28:** Final Learning Analytics Capability Model

The evaluation took a mixed-method approach: data was collected via pluralistic walkthroughs, a group discussion, and a survey. The model was evaluated on the concepts *effectiveness, usefulness*, and *completeness*. During the pluralistic walkthroughs, 26 practitioners from five different Dutch-speaking educational institutions used the Learning Analytics Capability Model to plan the implementation of learning analytics at their institution. The group discussion involved seven Dutch experts, experienced with learning analytics implementation. Also, the pluralistic walkthroughs involved group discussions at the end of each session. The survey was distributed among all participants (both practitioners and experts). Collectively, the collected data was used to evaluate the Learning Analytics Capability Model empirically. The model's effectiveness, i.e., its helpfulness to practitioners who want to implement learning analytics at their institution, was established during the pluralistic walkthroughs. When the model was used to enhance the implementation plan ('roadmap'), there was a 45% increase in capabilities included in the plans. The model's usefulness was established via group discussions and the survey's closed-ended questions. Participants commented that the model helps identify overlooked capabilities and discuss what steps to take next in implementation planning. The survey yielded the same results, although it indicates that the model's ease-of-use is borderline and might need improvement. Finally, the model's completeness was established via group discussions and the survey's open-ended question. The participants believe that the model is complete, i.e., it contains all elements necessary for successful learning analytics implementation. Nonetheless, both practitioners and experts mentioned that the model pays limited attention to the pedagogical use of learning analytics and the pedagogical interventions that follow learner data analysis and visualization. However, one must realize that the model aims to better understand what organizational capabilities to develop for successful learning analytics but not in what way to conduct successful learning analytics. After all, this depends on contextual factors like learning analytics goals, organizational strategy, pedagogical choices, and learning designs. This differs from institution to institution. The model recognizes the importance of these elements. That is, the model prescribes to consider contextual elements, to identify what changes need to be made to the current context, to support learning analytics with pedagogical theory, and to provide actionable recommendations to stakeholders. Detailing these elements is left to those close to the situation in which the learning analytics are going to be used. Nevertheless, highlighting the contextual and pedagogical influences might be improved in future research. A second, related improvement regards the performance monitoring of learning analytics-driven interventions. Although the model contains a capability called *Performance monitoring*, this mainly relates to the learning analytics system. From the ten underlying operational definitions, only one explicitly states to “establish qualitative and quantitative indicators of success”. While coding the roadmaps and pluralistic walkthrough transcripts, it became apparent that stakeholders believe it is important to measure the impact of learning analytics on learning. This element should be considered in future research as well.

## 8.2. Discussion

The Learning Analytics Capability Model provides an answer to the Ph.D. research’s main research question of what IS-capabilities are necessary for the successful adoption of learning analytics by Dutch higher educational institutions. With success, we mean that learning analytics benefit teachers and learners. Although their roles are different, both can take profit from learning analytics. For example, while teachers can use learning analytics to increase the effectiveness of their courses (Dyckhoff et al., 2012), learners can use it to support self-reflection and self-regulation (Wong & Li, 2020). The focus on teachers and learners aligns with the idea that learning analytics should be applied exclusively to the learning part of education (Long & Siemens, 2011). In this section, we discuss the relation between the research’s outcomes and the main research question.

### 8.2.1. A Model for learning analytics capabilities

The Learning Analytics Capability Model prescribes the development of 30 distinct capabilities that are grouped in five critical categories. Each category should be present and instantiated: when one category is missing, a higher educational institution cannot achieve successful learning analytics. The first category is *Data*. Learning analytics is about collecting and using learner data (Duval, 2012; Ifenthaler, 2015; LAK, 2011), so possessing and using the right data is fundamental to successful learning analytics. The second category is *Management*, which describes in what way learning analytics activities should be managed and governed. Governance is a common theme within Information Systems literature related to data analytics (de Bruin, 2009). Likewise, data governance and analytics governance need consideration when adopting learning analytics (Grochow, 2012). The third category is *People*. Many stakeholder groups exist in learning analytics research (Khalil & Ebner, 2015), and learning analytics adoption can only be achieved when individuals decide to use it in their work practices (Klein et al., 2019) or learning practices. Learners are at the center of learning analytics and must be involved in learning analytics design and implementation (Wise, 2014). The fourth category is *Technology*. Technology brings new opportunities and is one of the main drivers of learning analytics (Ferguson, 2012; Joksimović et al., 2019). For higher educational institutions, it is important to know what technologies to deploy to make learning analytics work. The fifth and final category is *Privacy & Ethics*. Learner data are personal data and thus subject to privacy laws and regulations (Engelfriet et al., 2017). Moreover, many ethical challenges must be overcome before learning analytics can be successfully and responsibly used (Slade & Prinsloo, 2013). The evaluation study presented in chapter 7 shows that the Learning Analytics Capability Model is effective, useful, and complete (Knobbout et al., 2021). With 818 operational definitions to operationalize the capabilities in practice, the model is extensive and provides a lot of information – see Appendix 2. This is necessary, as practical guidance on how to operationalize learning analytics at scale is absent from existing models (Broos et al., 2020). Thorough development of capabilities is required for successful learning analytics, as shown next.



### 8.2.2. Supporting learning analytics adoption by Dutch higher educational institutions

The case study presented in chapter 3 describes the problems encountered when implementing a learning analytics tool (Knobbout & van der Stappen, 2017). Mapping these problems on the Learning Analytics Capability Model, it becomes apparent that several capabilities were missing in the case organization – see Table 29. The data quality was low as not all activities were tracked. Therefore, the resulting graphs in the dashboard did not reflect what happened in reality. Also, only data from a single virtual learning environment were collected. This provided a one-dimensional perspective on the students' learning process, which can hamper learning analytics activities (Rienties et al., 2017). Regarding the model's category *People*, problems arose when teachers did not exactly know what activities to track. As a result, the visualizations could not be used to engage effective interventions. This confirms that training is an essential part of learning analytics (Rienties et al., 2018) Also, mainly better-performing students provided consent to collect data, which led to skewed descriptive statistics. Moreover, stakeholders from the IT department were not involved in time. This resulted in technical difficulties. Finally, the category *Technology* was directly affected when the connectivity between the case organization's virtual learning environment and the provider's online database and dashboard could not be established. Due to (the combination of) these problems, the learning analytics were not actively used for pedagogical interventions, in turn failing to positively impact learning. This case shows that many learning analytics capabilities need careful consideration before starting with analytical activities.

| **Issue encountered** (Knobbout & van der Stappen, 2017) | **Related capability from Learning Analytics Capability Model** |
|---|---|
| Only activities performed on one device were tracked | Data – Quality |
| Only VLE data was collected | Data – Sourcing & Integration |
| Collaborating students were not tracked | Data – Quality |
| Teachers did not know what activities to track | People – Combines Skills & Knowledge |
| Self-selection bias by students | People – Stakeholder Identification & Engagement |
| Stakeholders from IT department were not enough involved | People – Stakeholder Identification & Engagement |
| Connectivity between systems was established just hours before the experiment's start | Technology – Connectivity |

**Table 29:** Encountered Implementation Issues Plotted on Learning Analytics Capability Model.

The case study described in chapter 4 zooms in on the category *Data* (Knobbout et al., 2019). Data preprocessing, including data cleaning, is an important step in learning analytics (Conijn et al., 2016; Siemens, 2013). This is reflected in various capabilities from the Learning Analytics Capability Model. The capability *Quality* relates to the data's correctness and the cleaning process. Underlying operational definitions state that learning analytics must provide correct and accurate information. The case study showed that data cleaning directly affects the learning analytics outcomes and thus jeopardizes their correctness and accuracy.

The capability *Transparency* highlights the need to be transparent in what steps are taken during data cleaning, for what reason, and what the effects are on the outcomes. Chapter 4 recommends to be always transparent about the cleaning process along this line of thinking and describe what has happened and why. This resonates in the capability *Reporting*, which prescribes the need to justify and explain what has been done with the data. Inability to meet all these demands may lead to ineffective interventions and a decreased trust in learning analytics.

### 8.2.3. Benefiting teachers and learners

Learning analytics aims to improve learning and the environment in which this learning takes place (LAK, 2011). Data is collected, analyzed, visualized, and then used to perform interventions (Clow, 2012). According to Wise (2014), *pedagogical learning analytics interventions* involve instructional, studying or other components that directly and immediately impact teaching and learning processes. The research presented in chapter 2 shows that learning analytics are used by a variety of users and for different reasons (Knobbout & van der Stappen, 2020b). Learners can use their data to, for example, reflect on their learning activities and behavior. On the other hand, teachers can use learning analytics to predict performance and act upon it accordingly or improve assessment and feedback services. As a result, higher educational institutions use learning analytics in various ways and to achieve different goals. This was noticeable at the participating institutions to the pluralistic walkthroughs described in chapter 7 (Knobbout et al., 2021). One institution wants to use learning analytics to increase academic success and retention. However, in line with Gašević et al. (2015), we argue that measuring proxies such as course grades and quiz scores do not accurately capture the concept of learning. Therefore, other measures must be included as well. The classification scheme presented in chapter 2 supports this task and provides operational definitions to measure the impact of interventions on learning (Knobbout & van der Stappen, 2020b). Regardless of the goals higher educational institutions have with learning analytics, it is imperative to evaluate learning analytics' effects on learning (Knight et al., 2020; Larrabee Sønderlund et al., 2018). This is also captured in the Learning Analytics Capability Model. The capability *Performance Monitoring* is defined as "in what way the performance of analytical processes and applications are measured." Finally, the capability *Stakeholder identification & engagement* describes the early involvement of stakeholders. To design effective learning analytics, it is important involve teachers and learners. After all, they are the learning analytics' main beneficiaries (Prieto-Alvarez et al., 2019).

## 8.3. Contributions and implications

Design Science Research aims at achieving two goals: 1) add knowledge to the existing knowledge base and 2) design artifacts with the ability to solve problems in the environment (Hevner et al., 2004). We will discuss the contributions and implications of the Ph.D. research accordingly in the next sub-sections.

### 8.3.1. Scientific contributions

The Ph.D. research benefits scholars in the learning analytics research domain in multiple ways. First, in chapter 2, we have provided operational definitions to measure learning affected by pedagogical learning analytics interventions (Knobbout & van der Stappen, 2020b). As learning is affected by many aspects, the definitions are classified according to the teaching and learning model proposed by Biggs and Telfer (1987). Using the definitions to describe the effects of learning analytics on learning will lead to more evidence about the benefits of learning analytics on education. This is easier said than done as learner data are often incomplete and dirty, making analyses prone to misinterpretation and bias (Slade & Prinsloo, 2013). Nonetheless, standardizing these operational definitions allows to better compare the interventions' effect sizes and establish knowledge on what works and whatnot. This is an important outcome, as learning analytics' benefits are often not clear yet (Francis et al., 2020). By supporting learning analytics adoption, the Learning Analytics Capability Model allows collecting evidence whether or not learning analytics is beneficial for learning. A second contribution to the learning analytics research domain is that the Learning Analytics Capability Model identified some important capabilities missing from extant learning analytics models. In chapter 5, we combined multiple models and included knowledge from adjacent research domains. The analysis showed that certain capabilities are currently missing in learning analytics literature. The case study described in chapter 6 showed that the capabilities absent from existing learning analytics models were indeed found in practice. Therefore, our research outcome helps other scholars improve their work or supports the development of newer models.

Researchers for other analytics domains can also benefit from the Learning Analytics Capability Model. The model shows that the category *Privacy & Ethics* should not be overlooked. While conducting the systematic literature review presented in chapter 5, it became apparent that this category is missing from the literature on business analytics and big data analytics capabilities (Knobbout & van der Stappen, 2020a). As one of the practitioners to the pluralistic walkthroughs in chapter 7 remarked, the Learning Analytics Capability Model is also useful for other domains. Although the analyses' subjects vary, from an organizational point of view it often does not matter whether analytics are used to improve education or, for example, supply chains or healthcare.

Regarding the Information System research field, the research described in chapter 7 adds knowledge as it shows that pluralistic walkthroughs are an effective method to evaluate artifacts (Knobbout et al., 2021). In the Information Systems domain, there is often debate about what method to apply when evaluating an artifact (Venable et al., 2016; Winter, 2008). In line with Prat et al.'s (2015) suggestion to pragmatically design an evaluation method, we used pluralistic walkthroughs to evaluate our artifact in a naturalistic setting. Following our example, other scholars might want to consider this method for evaluation purposes as well. A second contribution to the Information System research field is that the Learning Analytics Capability Model is one of few models that are designed on Design Science Research principles (see Dawson et al. (2019) and Nguyen et al. (2020) for other examples). Moreover, the Learning Analytics Capability Model is the first model taking a resource-based perspective on learning analytics adoption to the best of our knowledge.

### 8.3.2. Practical contributions

Next to scientific contribution, this Ph.D. research benefits practitioners as well. With practice-based research in mind, Greven & Andriessen (2019) distinguish four types of impact pathways, i.e., knowledge development, product development, personal development, and systems development. While knowledge development is mainly covered by the scientific contributions (see section 8.3.1.), the other three impact forms can be regarded as practical contributions. We categorize our contributions accordingly.



From a product development point of view, the Learning Analytics Capability Model itself is a significant contribution to practice. It supports the adoption of learning analytics by European higher educational institutions and Dutch ones in particular. A model for the Dutch context is highly needed since most research regarding learning analytics adoption is focused on North America, the United Kingdom, and Australia (Ferguson & Clow, 2017; Yau & Ifenthaler, 2020). During the research, the model is adapted to the Dutch context by refining the model using empirical data from a Dutch university (chapter 6) and by having practitioners from Dutch(-speaking) (higher) educational institutions evaluate it (chapter 7). A major benefit of the Learning Analytics Capability Model is that it provides users with operational definitions that help develop practice capabilities – see Appendix 2. This supports program managers, policymakers, senior managers, et cetera, in their task of making strategic and actionable plans towards learning analytics adoption in their institution. Extant learning analytics models often miss these practical guidelines on how to operationalize capabilities (Broos et al., 2020). Towards this end, practical tooling is developed. To let practitioners work with the model, it is instantiated via the LACM tool and Excel tooling (chapter 7). By using these tools, higher educational institutions get insights into what capabilities are already present in the organization and which ones are yet missing. During future work, the tools might be improved and, for example, provide detailed advice on how to develop the absent capabilities.

Over the course of the Ph.D. research, results and insights are often shared with actors from the field, thereby contributing to their personal development. This was done at public events like the SURF Education Days (Knobbout et al., 2018; SURF, 2020) or practice-oriented conferences (Knobbout, 2019). Taking a hands-on approach, the presentations given at these events helped practitioners to take the next step towards learning analytics adoption at their institution. The recommendations provided in this thesis also implicitly add to personal development. The study presented in chapter 4 shows the effects of data cleaning on learning analytics outcomes. It is important to make users of learning analytics aware of the (unspoken) choices made during the data cleaning process. We provide recommendations on how to provide users insight into these choices and on how to improve the quality of data after they have been cleaned. Moreover, the classification scheme for affected learning (chapter 2) helps practitioners to be clear and transparent about the operational definitions they used in their efforts to improve learning via learning analytics.

From a system development perspective, local systems like educational institutes benefit from the Learning Analytics Capability Model as it helps to overcome challenges related to learning analytics adoption. The case organizations that participated in our evaluation study (chapter 7) are impacted by the model as it helped to structure their implementation plans. While the model's true practical value can only be shown via a longitudinal study with multiple organizations, the first reactions are positive. Initial results show that the Learning Analytics Capability Model leads to transformation in systems, thereby impacting the whole higher educational domain.

## 8.4. Limitations and directions for future research

The work described in this thesis has its limitations. This section summarizes these limitations and provides directions for future research.

### 8.4.1. Limitations

The first shortcoming of the model is that it is not fully implemented in practice. As time is a constraint during a Ph.D. research, the possibilities to apply the model to a case where a higher educational institution implemented learning analytics using the Learning Analytics Capability Model were limited. As a result, the model was only used in practice during pluralistic walkthroughs. To simulate the external environment, these pluralistic walkthroughs involved real users - policymakers, IT staff, institutional researchers, and teachers - from real institutions with real plans regarding learning analytics. During the pluralistic walkthroughs, users made an implementation roadmap with the model's help. Nonetheless, the effects of using the model during such sessions might differ from long-term application in practice. Future research should focus on its longitudinal application by higher educational institutions to better understand the model's use in practice.

A second, methodological shortcoming regards the studies described in chapters 3, 4, and 6. They rely on single-case studies. One might argue that multiple-case studies would have yielded better, more reliable results. Multiple-case studies indeed have some advantages over single-case studies - see Gustafsson (2017) for a detailed discussion on both methods. In general, evidence produced by multiple-case studies is strong and reliable (Baxter et al., 2008), is more likely to be confident and representative (Gerring, 2004), and allows for the comparison of similarities and differences between cases (Baxter et al., 2008; Yin, 2013). Nonetheless, there are reasons to choose single-case studies instead. Most noticeably, single-case studies can lead to high-quality theory as a result of a deeper understanding of the subject (Dyer & Wilkins, 1991), or might be used when a researcher wants to question theoretical relationships (Yin, 2013). Moreover, single-case studies can be conducted when the case presents an unusual case (Yin, 2013). Since not many institutions successfully apply learning analytics (Gašević et al., 2019), higher educational institutions that already use learning analytics provide these unusual cases. Such institutions are involved in this Ph.D. research. Finally, from a pragmatic point of view, single-case studies are cheaper and less time-consuming than multiple-case studies (Dyer & Wilkins, 1991; Gerring, 2004). Taking these arguments into consideration, we argue that single-cases studies are a suitable method.

The model is mainly focused on the organizational side of learning analytics. Nonetheless, Gašević et al. (2015) remind us not to forget that learning analytics are about learning. Learning analytics interacts in the middle space between learning sciences and educational research – both mature research fields that must not be overlooked when conducting learning analytics research (Knight et al., 2014). After all, learning analytics is about improving learning and intervention design is an important aspect (Wise, 2014). This was also remarked by the participants to the pluralistic walkthroughs described in chapter 7 (Knobbout et al., 2021). In the Learning Analytics Capability Model, the capability *Evidence-based & Theory-driven* highlights the importance to support learning analytics with (pedagogical) theory. However, as learning analytics as such is pedagogically neutral (Greller & Drachsler, 2012), the model does not describe what theory to follow. This depends on the context in which the model is applied. Considerations regarding the context are present in multiple capabilities.

A final limitation is that capabilities are not linked to the goals institutions have with learning analytics. The participating organizations in chapter 7 have different aims with learning analytics. Their goals may influence what capabilities are necessary. Imagine an institution that wishes to improve the learning environment by using learning analytics to enhance learning materials after the completion of each course. Due to the limited number of times data needs to be handled and analyzed, this institution might not need *Automation*. However, for an institution that wants to use learning analytics to track large numbers of students' online activities and behaviors simultaneously, the capability *Automation* is crucial to develop. The current research does not answer the question of what capabilities are necessary for what goals and context. To do so, we require a large number of higher educational institutions that have implemented learning analytics and monitor its performance. That way, the capabilities' effects on learning can be statistically researched. Such an approach is taken by, for example, Gupta & George (2016). However, this could not be done given the limited time available to conduct the entire Ph.D. research.

### 8.4.2. Directions for future research

At this moment, the Learning Analytics Capability Model is not fully implemented in practice yet. Future research should focus on its longitudinal application by higher educational institutions to better understand the model's use in practice. To support this task, additional materials might be developed. The Learning Analytics Capability Model tool provides a good starting point. This tool now only provides percentages to what degree capabilities are present at an institution and a list with the capabilities' operational definitions. Further development could focus on establishing formal maturity levels and provide detailed information on how to build the necessary capabilities. Also, to help higher educational institutions decide in what order to develop capabilities, the Learning Analytics Capability Model can be further refined into a process model. Process models are considered most useful to support learning analytics adoption by higher educational institutions (Broos et al., 2020; Colvin et al., 2017; Dawson et al., 2018). In chapter 7, preliminary results on what capabilities to develop in what phase of learning analytics adoption are provided (Knobbout et al., 2021). However, additional research towards this end is required.

Momentarily, it is unknown what the capabilities' effects on the learning analytics outcomes are. Some capabilities may contribute more to improved learning than others. For example, the ability to use data from multiple internal sources (*Sourcing & Integration*) can have a greater effect on improving learning processes than the capacity to use data from outside the institution (*External Environment*). We suggest statistically research the capabilities' effects on learning analytics performance. This helps higher educational institutions to effectively allocate resources and maximize the effect of learning analytics on learning processes, learning environments, and learning outcomes.

# References

Adrian, C., Abdullah, R., Atan, R., & Jusoh, Y. Y. (2018). Conceptual Model Development of Big Data Analytics Implementation Assessment Effect on Decision-Making. *International Journal of Interactive Multimedia and Artificial Intelligence*, *5*(1), 101–106.

Akçapinar, G. (2015). How automated feedback through text mining changes plagiaristic behavior in online assignments. *Computers and Education*, *87*, 123–130. https://doi.org/10.1016/j.compedu.2015.04.007

Akter, S., Wamba, S. F., Gunasekaran, A., Dubey, R., & Childe, S. J. (2016). How to improve firm performance using big data analytics capability and business strategy alignment? *International Journal of Production Economics*, *182*, 113–131.

Ali, L., Asadi, M., Gašević, D., Jovanović, J., & Hatala, M. (2013). Factors influencing beliefs for adoption of a learning analytics tool: An empirical study. *Computers & Education*, *62*, 130–148. https://doi.org/https://doi.org/10.1016/j.compedu.2012.10.023

Aljohani, N. R., Daud, A., Abbasi, R. A., Alowibdi, J. S., Basheri, M., & Aslam, M. A. (2019). An integrated framework for course adapted student learning analytics dashboard. *Computers in Human Behavior*, *92*, 679–690. https://doi.org/10.1016/j.chb.2018.03.035

Ambrosini, V., Bowman, C., & Collier, N. (2009). Dynamic capabilities: An exploration of how firms renew their resource base. *British Journal of Management*, *20*, S9–S24.

Amit, R., & Schoemaker, P. J. H. (1993). Strategic assets and organizational rent. *Strategic Management Journal*, *14*(1), 33–46. https://doi.org/10.1002/smj.4250140105

Arguedas, M., Daradoumis, T., & Xhafa, F. (2016). Analyzing the effects of emotion management on time and self-management in computer-based learning. *Computers in Human Behavior, 63*, 517–529. https://doi.org/10.1016/j.chb.2016.05.068

Arnold, K. E., Lonn, S., & Pistilli, M. D. (2014). An exercise in institutional reflection: The learning analytics readiness instrument (LARI). *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge*, 163–167.

Arnold, K. E., Lynch, G., Huston, D., Wong, L., Jorn, L., & Olsen, C. W. (2014). Building institutional capacities and competencies for systemic learning analytics initiatives. *ACM International Conference Proceeding Series*, 257–260. https://doi.org/10.1145/2567574.2567593

Aslan, S., Alyuz, N., Tanriover, C., Mete, S. E., Okur, E., D'Mello, S. K., & Esme, A. A. (2019). Investigating the Impact of a Real-time, Multimodal Student Engagement Analytics Technology in Authentic Classrooms. *Conference on Human Factors in Computing Systems - Proceedings*, 12. https://doi.org/10.1145/3290605.3300534

Avella, J. T., Kebritchi, M., Nunn, S., & Kanai, T. (2016). Learning analytics methods, benefits, and challenges in higher education: A systematic literature review. *Online Learning*, *20*(2).

Baker, R. S. (2019). Challenges for the future of educational data mining: The Baker learning analytics prizes. *Journal of Educational Data Mining (JEDM)*, *11*(1), 1–17.

Barneveld, A. Van, Arnold, K. E., & Campbell, J. P. (2012). Analytics in higher education: Establishing a common language. *EDUCAUSE Learning Initiative*, *1*, 1–11.

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, *17*(1), 99–120.

Baxter, P., Jack, S., & others. (2008). Qualitative case study methodology: Study design and implementation for novice researchers. *The Qualitative Report*, *13*(4), 544–559.

Beheshitha, S. S., Hatala, M., Gašević, D., & Joksimović, S. (2016). The role of achievement goal orientations when studying effect of learning analytics visualizations. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16*, 54–63. https://doi.org/10.1145/2883851.2883904

Ben David, Y., Segal, A., & Gal, Y. K. (2016). Sequencing educational content in classrooms using Bayesian knowledge tracing. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, 354–363.

Berg, Alan; Van den Bogaard, M.; Drachsler, H.; Filius, R.; Manderveld, J.; Schuwer, R. (2015). Grand Challenges Learning Analytics & Open En Online Onderwijs. Een verkenning. In *Surf*.

Berland, M., Davis, D., & Smith, C. P. (2015). AMOEBA: Designing for collaboration in computer science classrooms through live learning analytics. *International Journal of Computer-Supported Collaborative Learning*, *10*(4), 425–447. https://doi.org/10.1007/s11412-015-9217-z

Bharadwaj, A. S. (2000). A resource-based perspective on information technology capability and firm performance: an empirical investigation. *MIS Quarterly*, 169–196.

Bias, R. G. (1994). The pluralistic usability walkthrough: coordinated empathies. *Usability Inspection Methods*, 63–76.

Bichsel, J. (2012). *Analytics in higher education: Benefits, barriers, progress, and recommendations*. EDUCAUSE Center for Applied Research.

Biggs, J. B., & Telfer, R. (1987). *The process of learning*. McGraw-Hill/Appleton & Lange.

Bodily, R., & Verbert, K. (2017). Trends and issues in student-facing learning analytics reporting systems research. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, 309–318.

Boudreau, M.-C., Gefen, D., & Straub, D. W. (2001). Validation in information systems research: A state-of-the-art assessment. *MIS Quarterly*, 1–16.

Braungart, M., Braungart, R., & Gramet, P. (2007). Applying learning theories to healthcare practice. In *Psychological Learning Theories* (p. 40). Jones & Bartlett Learning. https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=2ahUKEwiMioaFjN3aAhXNmLQKHd_1Bz0QFjAAegQIABAs&url=http%3A%2F%2Fwww.jblearning.com%2Fsamples%2F0763751375%2Fchapter2.pdf&usg=AOvVaw3nZnvXOl-r7nHmiM-vlbFU

Brennan, R., Attard, J., & Helfert, M. (2018). Management of data value chains, a value monitoring capability maturity model. *20th International Conference on Enterprise Information Systems*, 12. https://doi.org/10.5220/0006684805730584

Brink, H., Richards, J., & Fetherolf, M. (2016). *Real-world machine learning*. Manning Publications Co.

Broos, T., Hilliger, I., Pérez-Sanagustín, M., Htun, N., Millecamp, M., Pesántez-Cabrera, P., Solano-Quinde, L., Siguenza-Guzman, L., Zuñiga-Prieto, M., Verbert, K., & De Laet, T. (2020). Coordinating learning analytics policymaking and implementation at scale. *British Journal of Educational Technology*, *51*(4), 938–954. https://doi.org/10.1111/bjet.12934

Bryman, A., & Bell, E. (2015). *Business research methods* (2nd ed.). Oxford University Press, USA.

Campbell, J. P., & Oblinger, D. G. (2007). Academic analytics. In *EDUCAUSE Quarterly*.

Carifio, J., & Perla, R. (2008). Resolving the 50-year debate around using and misusing Likert scales. *Medical Education*, *42*(12), 1150–1152.

Carvalho, J. Á. (2012). *Validation criteria for the outcomes of design research*.

Chen, B., Chang, Y. H., Ouyang, F., & Zhou, W. (2018). Fostering student engagement in online discussion through social learning analytics. *Internet and Higher Education*, *37*(April 2018), 21–30. https://doi.org/10.1016/j.iheduc.2017.12.002

Chen, D. Q., Preston, D. S., & Swink, M. (2015). How the use of big data analytics affects value creation in supply chain management. *Journal of Management Information Systems*, *32*(4), 4–39.

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, *36*(4), 1165–1188.

Cheng, H. C., & Liao, W. W. (2012). Establishing an lifelong learning environment using IOT and learning analytics. *International Conference on Advanced Communication Technology, ICACT*, 1178–1183.

Ciolacu, M., Tehrani, A. F., Binder, L., & Svasta, P. M. (2018). Education 4.0 - Artificial Intelligence Assisted Higher Education: Early recognition System with Machine Learning to support Students' Success. *2018 IEEE 24th International Symposium for Design and Technology in Electronic Packaging, SIITME 2018 - Proceedings*, *Ml*, 23–30. https://doi.org/10.1109/SIITME.2018.8599203

Clow, D. (2012). The learning analytics cycle: closing the loop effectively. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 134–138. https://doi.org/10.1145/2330601.2330636

Colvin, C., Dawson, S., Wade, A., & Gašević, D. (2017). Addressing the Challenges of Institutional Adoption. In C. Lang, G. Siemens, A. Wise, & D. Gašević (Eds.), *Handbook of Learning Analytics* (1st ed., pp. 281–289). https://doi.org/10.18608/hla17

Colvin, C., Rogers, T., Wade, A., Dawson, S., Gašević, D., Shum, S. B., & Fisher, J. (2015). Student retention and learning analytics: A snapshot of Australian practices and a framework for advancement. *Sydney: Australian Office for Learning and Teaching.*

Conijn, R., Snijders, C., Kleingeld, A., & Matzat, U. (2016). Predicting student performance from LMS data: A comparison of 17 blended courses using Moodle LMS. *IEEE Transactions on Learning Technologies*, *10*(1), 17–29.

Conijn, R., Snijders, C., Matzat, U., Kleingeld, A., & Nij Bijvank, W. (2016). *EXCTRA - EXploiting the Click-TRAil.*

Cooper, P. a. (1993). Paradigm Shifts in Designed Instruction- From Behaviorism to Cognitivism to Constructivism. *Educational Technology*, *35*(3), 12–19. https://www.learntechlib.org/p/170882/share/

Cosic, R., Shanks, S., & Maynard, G. (2015). A business analytics capability framework. *Australasian Journal of Information Systems*, *19*. https://doi.org/10.3127/ajis.v19i0.1150

Dahl, S. G., Allender, L., Kelley, T., & Adkins, R. (1995). Transitioning Software to the Windows Environment-Challenges and innovations. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *39*(18), 1224–1227.

Dahlberg, P. (2003). *Local mobility [PhD thesis]*. Göteborg University.

Davenport, T. H., & Harris, J. G. (2007). *Competing on analytics: The new science of winning*. Harvard Business Press.

Davis, D., Jivet, I., Kizilcec, R. F., Chen, G., Hauff, C., & Houben, G.-J. (2017). Follow the Successful Crowd: Raising MOOC Completion Rates through Social Comparison at Scale. *The Seventh International Learning Analytics & Knowledge Conference*, 454–463. https://doi.org/10.1145/3027385.3027411

Davis, F. D. (1985). *A technology acceptance model for empirically testing new end-user information systems: Theory and results*. Massachusetts Institute of Technology.

Dawson, S., Joksimovic, S., Poquet, O., & Siemens, G. (2019). Increasing the impact of learning analytics. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge (LAK)*, 446–455.

Dawson, S., Jovanovic, J., Gašević, D., Pardo, A., & AbelardoPardo, A. (2017). From prediction to impact: Evaluation of a learning analytics retention program. *Proceedings of the 7th International Learning Analytics & Knowledge Conference (LAK'17)*, 474–478. https://doi.org/10.1145/3027385.3027405

Dawson, S., Poquet, O., Colvin, C., Rogers, T., Pardo, A., & Gašević, D. (2018). Rethinking learning analytics adoption through complexity leadership theory. *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, 236–244.

Dawson, S., Poquet, O., Colvin, C., Rogers, T., Pardo, A., & Gašević, D. (2018). Rethinking learning analytics adoption through complexity leadership theory. *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, 236–244.

de Bruin, T. (2009). *Business process management : theory on progression and maturity.*

Del Blanco, Á., Serrano, Á., Freire, M., Martínez-Ortiz, I., & Ferníndez-Manjón, B. (2013). E-Learning standards and learning analytics. Can data collection be improved by using standard data models? *Global Engineering Education Conference (EDUCON), 2013 IEEE*, 1255–1261.

Diana, N., Eagle, M., Stamper, J., Grover, S., Bienkowski, M., & Basu, S. (2017). An instructor dashboard for real-time analytics in interactive programming assignments. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference on - LAK '17*, 272–279. https://doi.org/10.1145/3027385.3027441

Drachsler, H., & Greller, W. (2016). Privacy and analytics: it's a DELICATE issue a checklist for trusted learning analytics. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, 89–98.

Drachsler, H., Stoyanov, S., & Specht, M. (2014). The Impact of Learning Analytics on the Dutch Education System. *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge*, 158–162. https://doi.org/10.1145/2567574.2567617

Duval, E. (2012). *Learning Analytics and Educational Data Mining*. https://erikduval.wordpress.com/2012/01/30/learning-analytics-and-educational-data-mining/

Dyckhoff, A. L., Zielke, D., Bültmann, M., Chatti, M. A., & Schroeder, U. (2012). Design and Implementation of a Learning Analytics Toolkit for Teachers. *Educational Technology & Society*, *15*(3), 58–76.

Dyer, W. G., & Wilkins, A. L. (1991). Better stories, not better constructs, to generate better theory: A rejoinder to Eisenhardt. *Academy of Management Review*, *16*(3), 613–619.

Eisenhardt, K. M., & Martin, J. A. (2000). Dynamic capabilities: what are they? *Strategic Management Journal*, *21*(10–11), 1105–1121.

Emaus, T., Versendaal, J., Kloos, V., & Helms, R. (2010). Purchasing 2.0: An explorative study in the telecom sector on the potential of Web 2.0 in purchasing. *Proceedings of the 7th International Conference on Enterprise Systems, Accounting and Logistics (ICESAL 2010), Rhodes Island, Greece.*

Engelfriet, A., Manderveld, J., & Jeunink, E. (2017). *Handreiking: Learning analytics onder de Wet bescherming persoonsgegevens*. https://www.surf.nl/handreiking-learning-analytics-onder-de-wet-bescherming-persoonsgegevens

Faria, A.-M., Sorensen, N., Heppen, J., Bowdon, J., Taylor, S., Eisner, R., & Foster, S. (2017). Getting students on track for graduation: Impacts of the Early Warning Intervention and Monitoring System after one year At American Institutes for Research. In *Regional Educational Laboratory Midwest* (Issue April). http://ies.

Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, *4*(5–6), 304–317.

Ferguson, R., & Clow, D. (2017). Where is the evidence? A call to action for learning analytics. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference on - LAK '17*, 56–65. https://doi.org/10.1145/3027385.3027396

Ferguson, R., Clow, D., Macfadyen, L., Essa, A., Dawson, S., & Alexander, S. (2014). Setting learning analytics in context: overcoming the barriers to large-scale adoption. *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge*, 251–253.

Fidalgo-Blanco, Á., Sein-Echaluce, M. L., García-Peñalvo, F. J., & Conde, M. Á. (2015). Using Learning Analytics to improve teamwork assessment. *Computers in Human Behavior*, *47*, 149–156. https://doi.org/http://dx.doi.org/10.1016/j.chb.2014.11.050

Field, A. (2005). *Discovering statistics using SPSS* (2nd ed.). Sage Publications, Inc.

Foster, C., & Francis, P. (2019). A systematic review on the deployment and effectiveness of data analytics in higher education to improve student outcomes. *Assessment & Evaluation in Higher Education*, *45*(6), 822–841. https://doi.org/10.1080/02602938.2019.1696945

Francis, P., Broughan, C., Foster, C., & Wilson, C. (2020). Thinking critically about learning analytics, student outcomes, and equity of attainment. *Assessment & Evaluation in Higher Education*, *45*(6), 811–821. https://doi.org/10.1080/02602938.2019.1691975

Gašević, D., Dawson, S., & Jovanovic, J. (2016). Ethics and privacy as enablers of learning analytics. *Journal of Learning Analytics*, *3*(1), 1–4.

Gašević, D., Dawson, S., Rogers, T., & Gašević, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, *28*, 68–84.

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, *59*(1), 64–71. https://doi.org/10.1007/s11528-014-0822-x

Gašević, D., Mirriahi, N., & Dawson, S. (2014). Analytics of the effects of video use and instruction to support reflective learning. *Proceedins of the Fourth International Conference on Learning Analytics And Knowledge - LAK '14*, 123-132. https://doi.org/10.1145/2567574.2567590

Gašević, D., Tsai, Y.-S., Dawson, S., & Pardo, A. (2019). How do we start? An approach to learning analytics adoption in higher education. *The International Journal of Information and Learning Technology*, *36*(4), 342–353.

Gašević, D., Tsai, Y.-S., Dawson, S., & Pardo, A. (2019). How do we start? An approach to learning analytics adoption in higher education. *The International Journal of Information and Learning Technology*, *36*(4), 342–353.

Gerring, J. (2004). What is a case study and what is it good for? *American Political Science Review*, 341–354.

Gong, L., & Liu, Y. (2019). Design and Application of Intervention Model based on Learning Analytics under Blended Learning Environment. *International Conference on Innovative Education and Technology 2019*, 225–229.

Gong, L., Liu, Y., & Zhao, W. (2018). Using Learning Analytics to Promote Student Engagement and Achievement in Blended Learning : An Empirical Study. *International Conference on E-Education, E-Business and E-Technology 2018*, 2–7.

Grann, J., & Bushway, D. (2014). Competency map: visualizing student learning to promote student success. *Proceedins of the Fourth International Conference on Learning Analytics And Knowledge - LAK '14*, 168–172. https://doi.org/10.1145/2567574.2567622

Grant, R. M. (1991). The resource-based theory of competitive advantage: implications for strategy formulation. *California Management Review*, *33*(3), 114–135.

Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly*, *37*(2), 337–355.

Greller, W., & Drachsler, H. (2012). Translating Learning into Numbers: A Generic Framework for Learning Analytics. *Educational Technology & Society*, *15*(3), 42–57.

Greven, K., & Andriessen, D. (2019). Practice-based Research Impact Model for Evaluation: PRIME. *EAIR 41st Annual Forum*.

Groba, A. R., Barreiros, B. V., Lama, M., Gewerc, A., & Mucientes, M. (2015). Using a learning analytics tool for evaluation in self-regulated learning. *Proceedings - Frontiers in Education Conference, FIE*, *2015-Febru*(February), 2484–2491. https://doi.org/10.1109/FIE.2014.7044400

Grochow, J. (2012). *IT Infrastructure to Support Analytics: Laying the Groundwork for Institutional Analytics.*

Guarcello, M. A., Levine, R. A., Beemer, J., Frazee, J. P., Laumakis, M. A., & Schellenberg, S. A. (2017). Balancing Student Success: Assessing Supplemental Instruction Through Coarsened Exact Matching. *Technology, Knowledge and Learning*, *22*(3), 335–352. https://doi.org/10.1007/s10758-017-9317-0

Guillot, R., Seanosky, J., Guillot, I., Boulanger, D., Guillot, C., Kumar, V., Fraser, S. N., & Kinshuk. (2018). Assessing learning analytics systems impact by summative measures. *Proceedings - IEEE 18th International Conference on Advanced Learning Technologies, ICALT 2018*, 188–190. https://doi.org/10.1109/ICALT.2018.00051

Gupta, M., & George, J. F. (2016). Toward the development of a big data analytics capability. *Information & Management*, *53*(8), 1049–1064.

Gustafsson, J. (2017). *Single case studies vs. multiple case studies: A comparative study.* https://www.academia.edu/download/58755264/case_study_single_and_multiple_case.pdf

Harrison, S., Villano, R., Lynch, G., & Chen, G. (2016). Measuring financial implications of an early alert system. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16*, 241–248. https://doi.org/10.1145/2883851.2883923

Helfat, C. E., & Peteraf, M. A. (2003). The dynamic resource-based view: Capability lifecycles. *Strategic Management Journal*, *24*(10), 997–1010.

Herodotou, C., Rienties, B., Boroowa, A., Zdrahal, Z., & Hlosta, M. (2019). A large-scale implementation of predictive learning analytics in higher education: the teachers' role and perspective. *Educational Technology Research and Development*, *67*(5), 1273–1306. https://doi.org/10.1007/s11423-019-09685-0

Herodotou, C., Rienties, B., Boroowa, A., Zdrahal, Z., Hlosta, M., & Naydenova, G. (2017). Implementing predictive learning analytics on a large scale. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference on - LAK '17*, 267–271. https://doi.org/10.1145/3027385.3027397

Hevner, A. R. (2007). A three cycle view of design science research. *Scandinavian Journal of Information Systems*, *19*(2), 87–92.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, *28*(1), 75–105.

Hilliger, I., Ortiz-Rojas, M., Pesántez-Cabrera, P., Scheihing, E., Tsai, Y.-S., Muñoz-Merino, P. J., Broos, T., Whitelock-Wainwright, A., & Pérez-Sanagustín, M. (2020). Identifying needs for learning analytics adoption in Latin American universities: A mixed-methods approach. *The Internet and Higher Education*, *45*. https://doi.org/https://doi.org/10.1016/j.iheduc.2020.100726

Hirsch, D. D. (2014). The glass house effect: Big Data, the new oil, and the power of analogy. *Maine Law Review*, *66*(2), 373–395.

Hoel, T., Griffiths, D., & Chen, W. (2017). The influence of data protection and privacy frameworks on the design of learning analytics systems. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, 243–252.

Holman, C., Aguilar, S. J., Levick, A., Stern, J., Plummer, B., & Fishman, B. (2015). Planning for success: how students use a grade prediction tool to win their classes. *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*, 260–264.

HOP, Belleman, B., & Flink, E. (2020). *Wéér een nieuwe samenwerking in hoger onderwijs?* Website. https://www.delta.tudelft.nl/article/analyse-weer-een-nieuwe-samenwerking-hoger-onderwijs

Hubalovsky, S., Hubalovska, M., & Musilek, M. (2019). Assessment of the influence of adaptive E-learning on learning effectiveness of primary school pupils. *Computers in Human Behavior*, *92*, 691–705. https://doi.org/10.1016/j.chb.2018.05.033

Ifenthaler, D. (2015). Learning analytics. In J. M. Spector (Ed.), *The SAGE encyclopedia of educational technology* (pp. 447–451). Sage.

Ifenthaler, D., & Yau, J. Y.-K. (2019). Higher education stakeholders' views on learning analytics policy recommendations for supporting study success. *International Journal of Learning Analytics and Artificial Intelligence for Education (IJAI)*, *1*(1), 28–42.

Ifenthaler, D., & Yau, J. Y.-K. (2020). Utilising learning analytics to support study success in higher education: a systematic review. *Educational Technology Research and Development*, *68*, 1961–1990. https://doi.org/https://doi.org/10.1007/s11423-020-09788-z

Jayaprakash, S. M., Moody, E. W., Lauría, E. J. M., Regan, J. R., & Baron, J. D. (2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics*, *1*(1), 6–47. https://doi.org/10.18608/jla.2014.11.3

Jeon, I., & Song, K.-S. (2019). The Effect of Learning Analytics System towards Learner's Computational Thinking Capabilities. *International Conference on Computer and Automation Engineering 2019*, 12–16. https://doi.org/10.1145/3313991.3314017

Jing, X., & Tang, J. (2017). Guess You Like: Course Recommendations in MOOCs. *Proceedings of the International Conference on Web Intelligence - WI '17*, 783–789. https://doi.org/10.1145/3106426.3106478

Joksimović, S., Kovanović, V., & Dawson, S. (2019). The journey of learning analytics. *HERDSA Review of Higher Education*, *6*, 37–63.

Joksimović, S., Poquet, O., Kovanović, V., Dowell, N., Mills, C., Gašević, D., Dawson, S., Graesser, A. C., & Brooks, C. (2018). How do we model learning at scale? A systematic review of research on moocs. *Review of Educational Research*, *88*(1), 43–86. https://doi.org/10.3102/0034654317740335

Jovanović, J., Gašević, D., Pardo, A., Dawson, S., & Whitelock-Wainwright, A. (2019). Introducing Meaning to Clicks: Towards Traced-measures of Self-efficacy and Cognitive Load. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, 511–520. https://doi.org/10.1145/3303772.3303782

Khalil, M., & Ebner, M. (2015). Learning analytics: principles and constraints. *EdMedia+ Innovate Learning*, 1789–1799.

Khan, I., & Pardo, A. (2016). Data2U: Scalable real time student feedback in active learning environments. *LAK '16: Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, 249–253. https://doi.org/10.1145/2883851.2883911

Klein, C., Lester, J., Rangwala, H., & Johri, A. (2019). Learning analytics tools in higher education: Adoption at the intersection of institutional commitment and individual action. *The Review of Higher Education*, *42*(2), 565–593.

Klock, A. C. T., Ogawa, A. N., Gasparini, I., & Pimenta, M. S. (2018). Integration of learning analytics techniques and gamification: An experimental study. *Proceedings - IEEE 18th International Conference on Advanced Learning Technologies, ICALT 2018*, 133–137. https://doi.org/10.1109/ICALT.2018.00039

Knight, S., Gibson, A., & Shibani, A. (2020). Implementing learning analytics for learning impact: Taking tools to task. *The Internet and Higher Education*, *45*. https://doi.org/https://doi.org/10.1016/j.iheduc.2020.100729

Knight, S., Shum, S. B., & Littleton, K. (2014). Epistemology, assessment, pedagogy: where learning meets analytics in the middle space. *Journal of Learning Analytics*, *1*(2), 23–47.

Knobbout, J. H. (2019). *Onze studenten als meetobject: interessant of illegaal?* https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwiel8Cqv9_uAhUozoUKHUBEBs0QFjAAegQIAhAC&url=https%3A%2F%2Fhusite.nl%2Fveiliginformatiedelen%2Fwp-content%2Fuploads%2Fsites%2F199%2F2019%2F05%2FJustian-Knobbout-Hogesch

Knobbout, J. H., Everaert, H., & van der Stappen, E.J. (2019). From dirty data to multiple versions of truth: How different choices in data cleaning lead to different learning analytics outcomes. *BLED 2019 Proceedings*, 49–66. https://aisel.aisnet.org/bled2019/57

Knobbout, J. H., Jacobs, S., & Vermaas, K. (2018). *De 10 beste tips voor learning analytics.* https://www.slideshare.net/DeOnderwijsdagen/de-10-beste-tips-voor-learning-analytics-karianne-vermaas-justian-knobbout-owd18

Knobbout, J. H., & van der Stappen, E.J. (2017). A first step towards learning analytics: Implementing an experimental learning analytics tool. *30th Bled EConference: Digital Transformation - From Connecting Things to Transforming Our Lives, BLED 2017.* https://doi.org/10.18690/978-961-286-043-1.23

Knobbout, J. H., & van der Stappen, E.J. (2020a). A Capability Model for Learning Analytics Adoption: Identifying organizational capabilities from literature on big data analytics, business analytics, and learning analytics. *International Journal of Learning Analytics and Artificial Intelligence for Education (IJAI)*, *2*(1), 47–66.

Knobbout, J. H., & van der Stappen, E.J. (2020b). Where is the Learning in Learning Analytics? A Systematic Literature Review on the Operationalization of Learning-Related Constructs in the Evaluation of Learning Analytics Interventions. *IEEE Transactions on Learning Technologies*. https://doi.org/10.1109/TLT.2020.2999970

Knobbout, J. H., & van der Stappen, E.J. (2018). Where Is the Learning in Learning Analytics? *European Conference on Technology Enhanced Learning*, 88–100.

Knobbout, J. H., van der Stappen, E., & Versendaal, J. (2020). Refining the Learning Analytics Capability Model. *Proceedings of the 26th Americas Conference on Information Systems (AMCIS).*

Knobbout, J. H., van der Stappen, E., Versendaal, J., & van de Wetering, R. (2021). A Comprehensive Model to Support the Adoption of Learning Analytics: A Mixed-Method Approach. *Manuscript Submitted for Publication*.

Kobayashi, V. B., Mol, S., & Kismihok, G. (2014). Labour Market Driven Learning Analytics. *Journal of Learning Analytics*, *1*(3), 207–210. https://doi.org/10.18608/jla.2014.13.24

Kohavi, R., Rothleder, N. J., & Simoudis, E. (2002). Emerging trends in business analytics. *Communications of the ACM*, *45*(8), 45–48.

Kolb, D. A. (1984). Experiential Learning: Experience as The Source of Learning and Development. In *Prentice Hall, Inc.* (Issue 1984). Prentice Hall. https://doi.org/10.1016/B978-0-7506-7223-8.50017-4

Kumar, V., Boulanger, D., Seanosky, J., Panneerselvam, K., & Somasundaram, T. S. (2014). Competence analytics. *Journal of Computers in Education*, *1*(4), 251–270.

Kusters, R. J. G., & Versendaal, J. (2013). Horizontal collaborative e-purchasing for hospitals: IT for addressing collaborative purchasing impediments. *Journal of International Technology and Information Management*, *22*(1), 65–83.

Kwon, O., Lee, N., & Shin, B. (2014). Data quality management, data usage experience and acquisition intention of big data analytics. *International Journal of Information Management*, *34*(3), 387–394.

LAK. (2011). 1st International Conference on Learning Analytics and Knowledge. *International Conference on Learning Analytics and Knowledge*, 2011. https://tekri.athabascau.ca/analytics/

Larrabee Sønderlund, A., Hughes, E., & Smith, J. (2018). The efficacy of learning analytics interventions in higher education: A systematic review. *British Journal of Educational Technology*, *50*(5), 2594–2618. https://doi.org/10.1111/bjet.12720

Lauría, E. J. M., Moody, E. W., Jayaprakash, S. M., Jonnalagadda, N., & Baron, J. D. (2013). Open academic analytics initiative: initial research findings. *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, 150–154.

Learning Analytics Community Exchange. (9999). *LACE Evidence Hub*. Learning Analytics Community Exchange. http://evidence.laceproject.eu/proposition/25/a-learning/

Lim, L. A., Gentili, S., Pardo, A., Kovanović, V., Whitelock-Wainwright, A., Gašević, D., & Dawson, S. (2019). What changes, and for whom? A study of the impact of learning analytics-based process feedback in a large course. *Learning and Instruction*, *in press*. https://doi.org/10.1016/j.learninstruc.2019.04.003

Long, P., & Siemens, G. (2011). Penetrating the Fog: Analytics in Learning and Education. *EDUCAUSE Review*, *46*(5), 30–32. http://search.proquest.com.proxy.library.vanderbilt.edu/docview/964183308/13AF5BC47C138E29FF2/5?accountid=14816%5Cnhttps://login.proxy.library.vanderbilt.edu/login?url=http://search.proquest.com/docview/964183308/13AF5BC47C138E29FF2/5?accountid=14816

Lonn, S., Aguilar, S. J., & Teasley, S. D. (2015). Investigating student motivation in the context of a learning analytics intervention during a summer bridge program. *Computers in Human Behavior*, *47*, 90–97. https://doi.org/10.1016/j.chb.2014.07.013

Lonn, S., Aguilar, S., & Teasley, S. D. (2013). Issues, challenges, and lessons learned when scaling up a learning analytics intervention. *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, 235–239.

Lukyanenko, R., & Parsons, J. (2015). Guidelines for Establishing Instantiation Validity in IT Artifacts: A Survey of IS Research Instantiation Validity and Artifact Sampling View project Participatory Design of User-generated Content Systems View project. *Springer*, *9073*, 430–438. https://doi.org/10.1007/978-3-319-18714-3_35

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an “early warning system” for educators: A proof of concept. *Computers & Education*, *54*(2), 588–599.

Macfadyen, L. P., & Dawson, S. (2012). Numbers are not enough. Why e-learning analytics failed to inform an institutional strategic plan. *Journal of Educational Technology & Society*, *15*(3), 149–163.

Maimbo, H., & Pervan, G. (2005). Designing a case study protocol for application in IS research. *PACIS 2005 Proceedings*, 1281–1292.

Makadok, R. (2001). Toward a synthesis of the resource-based and dynamic-capability views of rent creation. *Strategic Management Journal*, *22*(5), 387–401.

Maldonado-Mahauad, J., Hilliger, I., Laet, T. De, Millecamp, M., Verbert, K., Ochoa, X., & Pérez-Sanagustín, M. (2018). The LALA Project: Building Capacity to Use Learning Analytics to Improve Higher Education in Latin America. *Companion Proceedings of the 8th International Learning Analytics & Knowledge Conference*, 630–637.

Manderveld, J. (2016). *Het learning analytics experiment*. https://communities.surf.nl/artikel/het-learning-analytics-experiment

Mangaroska, K., & Giannakos, M. (2019). Learning Analytics for Learning Design: A Systematic Literature Review of Analytics-Driven Design to Enhance Learning. *IEEE Transactions on Learning Technologies*, *12*(4), 516–534. https://doi.org/10.1109/TLT.2018.2868673

Mangaroska, K., Sharma, K., Giannakos, M., Trætteberg, H., & Dillenbourg, P. (2018). Gaze insights into debugging behavior using learner-centred analysis. *Proceedings of the 8th International Conference on Learning Analytics and Knowledge - LAK ’18*, 350–359. https://doi.org/10.1145/3170358.3170386

Manske, S., & Hoppe, H. U. (2016). The “Concept cloud”: Supporting collaborative knowledge construction based on semantic extraction from learner-generated artefacts. *Proceedings - IEEE 16th International Conference on Advanced Learning Technologies, ICALT 2016*, *July*, 302–306. https://doi.org/10.1109/ICALT.2016.123

March, S. T., & Smith, G. F. (1995). Design and natural science research on information technology. *Decision Support Systems*, *15*(4), 251–266.

Marcos-García, J. A., Martínez-Monés, A., & Dimitriadis, Y. (2015). DESPRO: A method based on roles to provide collaboration analysis support adapted to the participants in CSCL situations. *Computers and Education*, *82*, 335–353. https://doi.org/10.1016/j.compedu.2014.10.027

Martinez-Maldonado, R., Yacef, K., & Kay, J. (2015). TSCL: A conceptual model to inform understanding of collaborative learning processes at interactive tabletops. *International Journal of Human Computer Studies*, *83*, 62–82. https://doi.org/10.1016/j.ijhcs.2015.05.001

Mavroudi, A., Giannakos, M., & Krogstie, J. (2018). Supporting adaptive learning pathways through the use of learning analytics: developments, challenges and future opportunities. *Interactive Learning Environments*, *26*(2), 206–220. https://doi.org/10.1080/10494820.2017.1292531

McKenzie, W. A., Perini, E., Rohlf, V., Toukhsati, S., Conduit, R., & Sanson, G. (2013). A blended learning lecture delivery model for large and diverse undergraduate cohorts. *Computers and Education*, *64*, 116–126. https://doi.org/10.1016/j.compedu.2013.01.009

Meixner, B., Matusik, K., Grill, C., & Kosch, H. (2014). Towards an easy to use authoring tool for interactive non-linear video. *Multimedia Tools and Applications*, *70*(2), 1251–1276.

Melero, J., Harnandez-Leo, D., Jing Sun, & Santos, P. (2015). How was the activity? A visualization support for a case of location-based learning design. *British Journal of Educational Technology*, *46*(2), 317–329. http://eds.a.ebscohost.com.ezproxy.taylors.edu.my/eds/pdfviewer/pdfviewer?vid=5&sid=a3f9842a-73ed-456b-a821-151f4016d628%40sessionmgr4007

Michos, K., & Hernández-Leo, D. (2018). Supporting awareness in communities of learning design practice. *Computers in Human Behavior*, *85*, 255–270. https://doi.org/10.1016/j.chb.2018.04.008

Mingers, J. (2001). Combining IS research methods: towards a pluralist methodology. *Information Systems Research*, *12*(3), 240–259.

Nguyen, A., Tuunanen, T., Gardner, L., & Sheridan, D. (2020). Design principles for learning analytics information systems in higher education. *European Journal of Information Systems*, *0*(0), 1–28. https://doi.org/10.1080/0960085X.2020.1816144

Norris, D. M., & Baer, L. L. (2013). Building organizational capacity for analytics. In *Educause Learning Initiative*.

Nouri, J., Ebner, M., Ifenthaler, D., Saqr, M., Malmberg, J., Khalil, M., Bruun, J., Viberg, O., González, M. Á. C., Papamitsiou, Z., & others. (2019). Efforts in Europe for Data-Driven Improvement of Education--A Review of Learning Analytics Research in Seven Countries. *International Journal of Learning Analytics and Artificial Intelligence for Education (IJAI)*, *1*(1), 8–27.

Nunn, S., Avella, J. T., Kanai, T., & Kebritchi, M. (2016). Learning analytics methods, benefits, and challenges in higher education: A systematic literature review. *Online Learning*, *20*(2), 13–29.

Nussbaumer, A., Hillemann, E.-C., Gütl, C., & Albert, D. (2015). A Competence-based Service for Supporting Self-Regulated Learning in Virtual Environments. *Journal of Learning Analytics*, *2*(1), 101–133. https://doi.org/10.18608/jla.2015.21.6

Nyumba, T. O., Wilson, K., Derrick, C. J., & Mukherjee, N. (2018). The use of focus group discussion methodology: Insights from two decades of application in conservation. *Methods in Ecology and Evolution*, *9*(1), 20–32. https://doi.org/https://doi.org/10.1111/2041-210X.12860

Panadero, E., Klug, J., & Järvelä, S. (2016). Third wave of measurement in the self-regulated learning field: when measurement and intervention come hand in hand. *Scandinavian Journal of Educational Research*, *60*(6), 723–735. https://doi.org/10.1080/00313831.2015.1066436

Papamitsiou, Z., & Economides, A. A. (2014). Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review of Empirical Evidence. *Educational Technology & Society*, *17*(4), 49–64. http://www.ifets.info/journals/17_4/4.pdf

Papoušek, J., Stanislav, V., & Pelánek, R. (2016). Evaluation of an adaptive practice system for learning geography facts. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16*, 134–142. https://doi.org/10.1145/2883851.2883884

Pappas, I. O., Mikalef, P., Giannakos, M. N., Krogstie, J., & Lekakos, G. (2018). Big data and business analytics ecosystems: paving the way towards digital transformation and sustainable societies. *Information Systems and E-Business Management*, *16*(3), 479–491. https://doi.org/10.1007/s10257-018-0377-z

Pardo, A., & Kloos, C. D. (2011). Stepping out of the box: towards analytics outside the learning management system. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*, 163–167.

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, *45*(3), 438–450.

Paré, G., Trudel, M.-C., Jaana, M., & Kitsiou, S. (2015). Synthesizing information systems knowledge: A typology of literature reviews. *Information & Management*, *52*(2), 183–199.

Peffers, K., Rothenberger, M., Tuunanen, T., & Vaezi, R. (2012). Design science research evaluation. *International Conference on Design Science Research in Information Systems*, 398–410.

Pérez-Sanagustín, M., Hilliger, I., Maldonado, J., Pérez, R., Ramírez, L., Muñoz-Merino, P. J., Tsai, Y., Ortiz, M., Broos, T., Zúñiga-Prieto, M., Sheihing, E., & Whitelock-Wainright, A. (2019). *LALA: Building Capacity to Use Learning Analytics to Improve Higher Education in Latin America*. https://www.lalaproject.org/wp-content/uploads/2019/04/LALA_framework_English.pdf

Perikos, I., Grivokostopoulou, F., & Hatzilygeroudis, I. (2017). Assistance and Feedback Mechanism in an Intelligent Tutoring System for Teaching Conversion of Natural Language into Logic. *International Journal of Artificial Intelligence in Education*, *27*(3), 475–514. https://doi.org/10.1007/s40593-017-0139-y

Picciano, A. G. (2014). Big data and learning analytics in blended learning environments: Benefits and concerns. *International Journal of Interactive Multimedia and Artificial Intelligence*, *2*(7), 35–43.

Pistilli, M. D., & Arnold, K. E. (2012). Course signals at Purdue: using learning analytics to increase student success. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, *May*, 2–5. https://doi.org/10.1145/2330601.2330666

Prakash, B. R., Hanumanthappa, M., & Kavitha, V. (2014). Big data in educational data mining and learning analytics. *Int. J. Innov. Res. Comput. Commun. Eng*, *2*(12), 7515–7520.

Prat, N., Comyn-Wattiau, I., & Akoka, J. (2015). A taxonomy of evaluation methods for information systems artifacts. *Journal of Management Information Systems*, *32*(3), 229–267.

Pries-Heje, J., Baskerville, R., & Venable, J. (2008). Strategies for Design Science Research Evaluation. In W. Golden, T. Acton, kieran Conboy, & V. K. Tuunainen (Eds.), *Conference Proceedings, 16th European Conference on Information Systems*. National University of Ireland.

Prieto-Alvarez, C. G., Martinez-Maldonado, R., & Anderson, T. (2019). Co-designing learning analytics tools with learners. In J. M. Lodge, J. C. Horvath, & L. Corrin (Eds.), *Learning Analytics in the Classroom: Translating Learning Analytics for Teachers*. Routledge.

Ren, S. J., Wamba, S. F., Akter, S., Dubey, R., & Childe, S. J. (2017). Modelling quality dynamics, business value and firm performance in a big data analytics environment. *International Journal of Production Research*, *55*(17), 5011–5026.

Reschly, A. L., & Christenson, S. L. (2012). Jingle, jangle, and conceptual haziness: Evolution and future directions of the engagement construct. In *Handbook of Research on Student Engagement* (pp. 3–19). Springer US. https://doi.org/10.1007/978-1-4614-2018-7_1

Rienties, B., Cross, S., & Zdrahal, Z. (2017). Implementing a Learning Analytics Intervention and Evaluation Framework: What Works? In K. D. B. (Ed.), *Big Data and Learning Analytics in Higher Education* (pp. 147–166). Springer International Publishing. https://doi.org/10.1007/978-3-319-06520-5_10

Rienties, B., Herodotou, C., Olney, T., Schencks, M., & Boroowa, A. (2018). Making Sense of Learning Analytics Dashboards: A Technology Acceptance Perspective of 95 Teachers. *International Review of Research in Open and Distributed Learning*, *19*(5). https://doi.org/10.19173/irrodl.v19i5.3493

Rienties, B., Toetenel, L., & Bryan, A. (2015). Scaling up learning design: impact of learning design activities on lms behavior and performance. *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*, 315–319.

Riihiaho, S. (2002). The pluralistic usability walk-through method. *Ergonomics in Design*, *10*(3), 23–27.

Rödle, W., Wimmer, S., Zahn, J., Prokosch, H.-U., Hinkes, B., Neubert, A., Rascher, W., Kraus, S., Toddenroth, D., & Sedlmayr, B. (2019). User-centered development of an online platform for drug dosing recommendations in pediatrics. *Applied Clinical Informatics*, *10*(04), 570–579.

Rome, M. (2011). *Best Practices in Student Learning and Assessment: Creating and Implementing Effective Assessment for NYU Schools, Departments* (p. 3). https://www.coursehero.com/file/24154206/Best-Practices-in-Student-Learning-Assessment-2011-02-17doc/

Romero-Zaldivar, V.-A., Pardo, A., Burgos, D., & Kloos, C. D. (2012). Monitoring student progress using virtual appliances: A case study. *Computers & Education*, *58*(4), 1058–1067.

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *10*(3). https://doi.org/10.1002/widm.1355

Rosmansyah, Y., Kartikasari, N., & Wuryandari, A. I. (2018). A learning analytics tool for monitoring and improving students' learning process. *Proceedings of the 2017 6th International Conference on Electrical Engineering and Informatics: Sustainable Society Through Digital Innovation, ICEEI 2017, 2017-Novem*, 1–5. https://doi.org/10.1109/ICEEI.2017.8312462

Rubel, A., & Jones, K. M. L. (2016). Student privacy in learning analytics: An information ethics perspective. *The Information Society*, *32*(2), 143–159.

Ruiz-Calleja, A., Prieto, L. P., Ley, T., Rodríguez-Triana, M. J., & Dennerlein, S. (2017). Learning analytics for professional and workplace learning: A literature review. *European Conference on Technology Enhanced Learning (EC-TEL 2017)*, *10474 LNCS*, 164–178. https://doi.org/10.1007/978-3-319-66610-5_13

Runeson, P., & Höst, M. (2009). Guidelines for conducting and reporting case study research in software engineering. *Empirical Software Engineering*, *14*(2), 131.

Schwendimann, B. A., Rodriguez-Triana, M. J., Vozniuk, A., Prieto, L. P., Boroujeni, M. S., Holzer, A., Gillet, D., & Dillenbourg, P. (2017). Perceiving learning at a glance: A systematic literature review of learning dashboard research. *IEEE Transactions on Learning Technologies*, *10*(1), 30–41. https://doi.org/10.1109/TLT.2016.2599522

Sclater, N. (2017). *Learning analytics explained*. Routledge. https://www.routledge.com/Learning-Analytics-Explained-1st-Edition/Sclater/p/book/9781138931732

Sclater, N., & Bailey, P. (2015). *Code of practice for learning analytics*. Jisc. https://www.jisc.ac.uk/guides/code-of-practice-for-learning-analytics

Seanosky, J., Guillot, I., Boulanger, D., Guillot, R., Guillot, C., Kumar, V., Fraser, S. N., Kinshuk, Aljojo, N., & Munshi, A. (2017). Real-Time Visual Feedback: A Study in Coding Analytics. *Proceedings - IEEE 17th International Conference on Advanced Learning Technologies, ICALT 2017*, 264–266. https://doi.org/10.1109/ICALT.2017.38

Sharma, K., Alavi, H. S., Jermann, P., & Dillenbourg, P. (2016). A gaze-based learning analytics model: in-video visual feedback to improve learner's attention in MOOCs. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16*, 417–421. https://doi.org/10.1145/2883851.2883902

Shimada, A., & Konomi, S. (2017). A Lecture Supporting System Based on Real-Time Learning Analytics. *International Association for Development of the Information Society*, *CELDA*, 197–204.

Siadaty, M., Gašević, D., & Hatala, M. (2016a). Measuring the impact of technological scaffolding interventions on micro-level processes of self-regulated workplace learning. *Computers in Human Behavior*, *59*(June), 469–482. https://doi.org/10.1016/j.chb.2016.02.025

Siadaty, M., Gašević, D., & Hatala, M. (2016b). Associations between technological scaffolding and micro-level processes of self-regulated learning: A workplace study. *Computers in Human Behavior*, *55*, 1007–1019. https://doi.org/10.1016/J.CHB.2015.10.035

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, *57*(10), 1380–1400.

Siemens, G. (2012). Learning analytics: envisioning a research discipline and a domain of practice. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 4–8.

Siemens, G., & Baker, R. S. J. d. (2012). Learning Analytics and Educational Data Mining: Towards Communication and Collaboration. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 252–254. https://doi.org/10.1145/2330601.2330661

Siemens, G., Dawson, S., & Lynch, G. (2013). Improving the quality and productivity of the higher education sector. In *Policy and Strategy for Systems-Level Deployment of Learning Analytics.*

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, *57*(10), 1510–1529.

Smolin, D., & Butakov, S. (2012). Applying Artificial Intelligence to the Educational Data: An Example of Syllabus Quality Analysis. *Proceedings of the Second International Conference on Learning Analytics and Knowledge - LAK '12*, *May*, 164–169. https://doi.org/10.1145/2330601.2330644

Strauss, A., & Corbin, J. (1990). Open coding. *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*, *2*(1990), 101–121.

SURF. (2020). *Vlootshow SURF Onderwijsdagen 2020*. https://www.surf.nl/vlootshow-surf-onderwijsdagen-2020/studiedata

Suthers, D., & Verbert, K. (2013). Learning analytics as a middle space. *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, 1–4.

Swidan, O., Prusak, N., Livny, A., Palatnik, A., & Schwarz, B. (2019). Fostering teachers' understanding of progression of multiple groups towards the orchestration of conceptual learning. *Unterrichtswissenschaft*, *47*(2), 159–176. https://doi.org/10.1007/s42010-019-00050-x

Syed, M., Anggara, T., Lanski, A., Duan, X., Ambrose, G. A., & Chawla, N. V. (2019). Integrated Closed-loop Learning Analytics Scheme in a First Year Experience Course. *The 9th International Learning Analytics & Knowledge Conference (LAK19)*, 521–530. https://doi.org/10.1145/3303772.3303803

Tabuenca, B., Kalz, M., Drachsler, H., & Specht, M. (2015). Time will tell: The role of mobile learning analytics in self-regulated learning. *Computers and Education*, *89*, 53–74. https://doi.org/10.1016/j.compedu.2015.08.004

Tai, J. H. M., Bellingham, R., Lang, J., & Dawson, P. (2019). Student perspectives of engagement in learning in contemporary and digital contexts. *Higher Education Research and Development*, *38*(5), 1075–1089. https://doi.org/10.1080/07294360.2019.1598338

Teaching Commons. (2019). *Direct Versus Indirect Assessment of Student Learning.* DePaul University. https://resources.depaul.edu/teaching-commons/teaching-guides/feedback-grading/Pages/direct-assessment.aspx

Tempelaar, D. T., Rienties, B., & Giesbers, B. (2015). In search for the most informative data for feedback generation: Learning analytics in a data-rich context. *Computers in Human Behavior*, *47*, 157–167. https://doi.org/10.1016/j.chb.2014.05.038

Thorvald, P., Lindblom, J., & Schmitz, S. (2015). Modified pluralistic walkthrough for method evaluation in manufacturing. *Procedia Manufacturing*, *3*, 5139–5146.

Thurmond, V. A. (2001). The point of triangulation. *Journal of Nursing Scholarship*, *33*(3), 253–258.

Tsai, Y.-S., & Gašević, D. (2017a). *The state of learning analytics in Europe.*

Tsai, Y.-S., & Gašević, D. (2017b). Learning analytics in higher education--challenges and policies: a review of eight learning analytics policies. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, 233–242.

Tsai, Y.-S., Moreno-Marcos, P. M., Tammets, K., Kollom, K., & Gašević, D. (2018). SHEILA policy framework: informing institutional strategies and policy processes of learning analytics. *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, 320–329.

Tsai, Y.-S., Poquet, O., Gašević, D., Dawson, S., & Pardo, A. (2019). Complexity leadership in learning analytics: Drivers, challenges and opportunities. *British Journal of Educational Technology*, *50*(6), 2839–2854.

van der Spek, E. (2018). *Samen aanjagen van vernieuwing: SURF meerjarenagenda 2019-2022*. https://www.surf.nl/media/2319

van Klaveren, C., Vonk, S., & Cornelisz, I. (2017). The effect of adaptive versus static practicing on student learning - evidence from a randomized field experiment. *Economics of Education Review*, *58*, 175–187. https://doi.org/10.1016/j.econedurev.2017.04.003

Van Leeuwen, A., Janssen, J., Erkens, G., & Brekelmans, M. (2014). Supporting teachers in guiding collaborating students: Effects of learning analytics in CSCL. *Computers and Education*, *79*, 28–39. https://doi.org/10.1016/j.compedu.2014.07.007

Van Leeuwen, A., Janssen, J., Erkens, G., & Brekelmans, M. (2015). Teacher regulation of cognitive activities during student collaboration: Effects of learning analytics. *Computers and Education*, *90*, 80–94. https://doi.org/10.1016/j.compedu.2015.09.006

van Leeuwen, A., Rummel, N., & van Gog, T. (2019). What information should CSCL teacher dashboards provide to help teachers interpret CSCL situations? *International Journal of Computer-Supported Collaborative Learning*, *14*(3), 261–289. https://doi.org/10.1007/s11412-019-09299-x

VanderPlas, J. (2016). *Python data science handbook: Essential tools for working with data* (1st ed.). " O'Reilly Media, Inc."

Vargas, H., Heradio, R., Chacon, J., De La Torre, L., Farias, G., Galan, D., & Dormido, S. (2019). Automated assessment and monitoring support for competency-based courses. *IEEE Access*, *7*, 41043–41051. https://doi.org/10.1109/ACCESS.2019.2908160

Venable, J. (2010). Design science research post hevner et al.: criteria, standards, guidelines, and expectations. *International Conference on Design Science Research in Information Systems*, 109–123.

Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: a Framework for Evaluation in Design Science Research. *European Journal of Information Systems*, *25*(1), 77–89. https://doi.org/10.1057/ejis.2014.36

Venkatesh, V., Brown, S. A., & Bala, H. (2013). Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems. *MIS Quarterly*, *37*(1), 21–54.

Vereniging van Universiteiten, Vereniging Hogescholen, & SURF. (2017). *Versnellingsagenda voor onderwijsinnovatie*. https://versnellingsplan.nl/wp-content/uploads/2019/11/Versnellingsagenda-2017.pdf

Vermunt, J. D., Ilie, S., & Vignoles, A. (2018). Building the foundations for measuring learning gain in higher education: A conceptual framework and measurement instrument. *Higher Education Pedagogies*, *3*(1), 266–301. https://doi.org/10.1080/23752696.2018.1484672

Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The current landscape of learning analytics in higher education. *Computers in Human Behavior*, *89*, 98–110. https://doi.org/10.1016/j.chb.2018.07.027

Wade, M., & Hulland, J. (2004). The resource-based view and information systems research: Review, extension, and suggestions for future research. *MIS Quarterly*, *28*(1), 107–142.

Wang, Y., & Byrd, T. A. (2017). Business analytics-enabled decision-making effectiveness through knowledge absorptive capacity in health care. *Journal of Knowledge Management*, *21*(3), 517–539.

Wang, Y., & Hajli, N. (2017). Exploring the path to big data analytics success in healthcare. *Journal of Business Research*, *70*, 287–299.

Wang, Y., Kung, L., Wang, W. Y. C., & Cegielski, C. G. (2018). An integrated big data analytics-enabled transformation model: Application to health care. *Information & Management*, *55*(1), 64–79.

Whale, S., Valenzuela, F. R., & Fisher, J. (2013). Implementing timely interventions to improve students' learning experience*. *30th Annual Conference on Australian Society for Computers in Learning in Tertiary Education, ASCILITE 2013*, 908–912.

Whitelock, D., Twiner, A., Richardson, J. T. E., Field, D., & Pulman, S. (2015). OpenEssayist: a supply and demand learning analytics tool for drafting academic essays. *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge - LAK '15*, 208–212. https://doi.org/10.1145/2723576.2723599

Winter, R. (2008). Design science research in Europe. *European Journal of Information Systems*, *17*(5), 470–475. https://doi.org/10.1057/ejis.2008.44

Wise, A. F. (2019). Learning Analytics: Using Data-Informed Decision-Making to Improve Teaching and Learning. In O. Adesope & A. Rud (Eds.), *Contemporary Technologies in Education* (pp. 119–143). Palgrave Macmillan. https://doi.org/10.1007/978-3-319-89680-9

Wise, A. F. (2014). Designing pedagogical interventions to support student use of learning analytics. *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge International Learning Analytics & Knowledge Conference on - LAK '14*, 203–211. https://doi.org/10.1145/2567574.2567588

Wise, A. F., & Vytasek, J. (2017). Learning analytics implementation design. *Handbook of Learning Analytics*, *1*.

Wise, A. F., Vytasek, J. M., Hausknecht, S., & Zhao, Y. (2016). Developing learning analytics design knowledge in the "middle space": The student tuning model and align design framework for learning analytics use. *Online Learning*, *20*(2), 28.

Wong, B. T. ming, & Li, K. C. (2020). A review of learning analytics intervention in higher education (2011–2018). *Journal of Computers in Education*, *7*(1), 7–28. https://doi.org/10.1007/s40692-019-00143-7

Xiong, X., Wang, Y., & Beck, J. B. (2015). Improving students' long-term retention performance. *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge - LAK '15*, 325–329. https://doi.org/10.1145/2723576.2723636

Yamada, M., Kitamura, S., Matsukawa, H., Misono, T., Kitani, N., & Yamauchi, Y. (2014). Collaborative filtering for expansion of learner's background knowledge in online language learning: does "top-down" processing improve vocabulary proficiency? *Educational Technology Research and Development*, *62*(5), 529–553. https://doi.org/10.1007/s11423-014-9344-7

Yau, J. Y.-K., & Ifenthaler, D. (2020). Reflections on different learning analytics indicators for supporting study success. *International Journal of Learning Analytics and Artificial Intelligence for Education*, *2*(2), 4–23. https://doi.org/10.3991/ijai.v2i2.15639

Yin, R. K. (2013). *Case study research: Design and methods*. Sage publications.

Zhang, J., Zhang, X., Jiang, S., de Pablos, P., & Sun, Y. (2018). Mapping the study of learning analytics in higher education. *Behaviour & Information Technology*, *37*(10–11), 1142–1155.

Zimmerman, B. J. (1989). A Social Cognitive View of Self-Regulated Academic Learning. *Journal of Educational Psychology*, *81*(3), 329–339. https://doi.org/10.1037/0022-0663.81.3.329

# Addendum

To improve the consistency and coherence of this thesis, the following changes were made to the original published articles:

- Typos and spelling mistakes were corrected.
- The article presented in chapter 7 was originally published using British-English spelling. This is changed to American-English, so it is consistent with the other chapters.
- In all articles, the figure and the table captions have been updated so they are consistent with the overall thesis.

# Summary

Driven by the digitalization of learning, the phrase "Data is the new oil" becomes true for the educational domain as well. While learning, students leave digital traces behind. Information systems allow using these traces to enhance learning. *Learning analytics* is the process of measuring, collecting, analyzing, and reporting on learner data to understand and optimize learning and the environment in which learning takes place. In the past decade, interest in this type of data analytics by educational institutions rapidly grew. Much of the initial research was conducted from a technological and data science perspective. However, nowadays the pedagogical and organizational sides are equally important. That is, higher educational institutions want to know what the effects of learning analytics on learning are and in what way learning analytics can be adopted at scale. To support adoption, various learning analytics models are developed. These models help institutions to start using learning analytics. Nonetheless, the extant models have their shortcomings. For example, not all models are empirically validated, they sometimes miss practical guidance, it remains unclear in what way to operationalize critical dimensions, and several models mainly focus on institutions in Angle-Saxon countries.

The Ph.D. research at hand fills the existing knowledge gaps and helps higher educational institutions to adopt learning analytics at scale by taking a resource-based perspective. Central to the resource-based view are resources and capabilities. Resources involve data, technology, people, funding et cetera. Capabilities describe in what way these resources should interact to benefit an organization. In contrast to resources, capabilities are non-transferable and must be developed by an organization itself. Consequently, higher educational institutions that want to adopt learning analytics need to develop learning analytics capabilities specific to their context. Without the right capabilities, learner data cannot be turned into insights and improvements to learning. This raises the question of what capabilities are important to learning analytics and in what way they must be developed in practice.

This Ph.D. research provides an answer to the main research question: "*Which Information System-related capabilities for learning analytics benefit teachers and learners in Dutch higher educational institutions?*" To answer this question, a capability model for learning analytics was developed by applying Design Science Research principles. A multimethod approach was utilized. This thesis comprises four interrelated parts. The first part describes the research's motivation and

in what way the effects of learning analytics on learning can be measured. The second part demonstrates the importance of having well-developed learning analytics capabilities. Next, the Learning Analytics Capability Model is developed and refined in the third part. The fourth and final part relates to the model's evaluation and discusses the research's outcomes.

Higher educational institutions that want to use learning analytics face many challenges. An exploratory case study showed that the lack of capabilities regarding data, technology, and stakeholders can pose serious issues while adopting a learning analytics tool. Another case study highlights the need to carefully consider the steps that must be taken during the data cleaning process. Data cleaning can affect the analyses' outcomes and thus the sequential learning analytics interventions. Therefore, stakeholders must be informed on what steps were taken, why they were taken, and what their consequences are. When learning analytics is adopted and used in practice, measuring the effects of learning analytics on learning help to determine the benefits to the stakeholders and the learning processes. According to learning theory, the measured effects can relate to three categories: the learning environment, learning process, and learning outcomes. Based on 62 key studies found during a systematic literature review, the categories were extended with eleven subcategories. The review's results help higher educational institutions systematically describe the effectiveness of learning analytics. In turn, this stimulates learning analytics adoption by other institutions.

Research towards resource-based capabilities for learning analytics is nascent. Therefore, knowledge from more mature research fields such as business analytics and big data analytics were exapted to the learning analytics domain during a systematic literature review. By analyzing ten key studies on business analytics and big data analytics, four capability categories prove important: Data, Management, People, and Technology. However, six additional key studies from the learning analytics domain provide a fifth category that must not be overlooked: Privacy & Ethics. Based on the systematic literature review's outcomes, a theoretical Learning Analytics Capability Model with 34 capabilities was developed. These capabilities are classified into the five categories. To help to operationalize the model in practice, the model comprises 461 operational definitions that support users to develop the capabilities in their institution. Next, a case study was conducted at an institution mature in the use of student data. The number of operationalizations rose to 818. Also, seven capabilities were merged, three capabilities were renamed, and all capabilities' definitions were improved.

During the research's final stage, the Learning Analytics Capability Model was evaluated with real users during five pluralistic walkthroughs. In total, 26 practitioners from five educational institutions used the model to enhance an implementation roadmap they made for their institution. An additional evaluation was done by having a group discussion with seven learning analytics experts and by surveying both the practitioners and experts. The evaluation study confirmed the model's effectiveness, usefulness, and completeness.

The Ph.D. research leads to several positive outcomes. From a practical perspective, the Learning Analytics Capability Model provides a model for Dutch higher educational institutions supporting stakeholders in their task of planning the adoption of learning analytics. Furthermore, it provides measures to assess the impact of learning analytics on learning, the benefits of learning analytics can be identified better. This helps to mature the learning analytics domain. Also, the model gives clear recommendations on the data clearing process. Not only practitioners but academia also benefits from this Ph.D. research. It extends existing knowledge on what capabilities are important to learning analytics adoption. It is important to note that the Learning Analytics Capability Model does not try to replace existing models. However, it continues the good work and overcomes several shortcomings of these extant models. Not only does the Learning Analytics Capability Model combines what is already known about learning analytics capabilities, but it also describes in what way the capabilities can be operationalized. Future research should aim to further establish in what order capabilities must be developed and identify the effect sizes of the various capabilities on the beneficial outcomes of learning analytics. Moreover, this research shows that Design Science Research and the resource-based view provide valuable information to the learning analytics domain and pluralistic walkthroughs can be used as a method for ex-post evaluation.

## Samenvatting

Gedreven door digitalisering wordt de uitspaak "data is de nieuwe olie" ook waarheid voor het educatieve domein. Tijdens het leren laten studenten digitale sporen achter. Informatiesystemen zijn in staat deze sporen te gebruiken om het leren te verbeteren. *Learning analytics* is het meten, verzamelen, analyseren en rapporteren van studentdata om het leren te begrijpen en te optimaliseren, evenals de leeromgeving waarin dit leren plaatsvind. Interesse in deze vorm van data-analyse is het afgelopen decennium flink gegroeid. Veel van het initiële onderzoek werd uitgevoerd vanuit technisch- en datawetenschappelijk perspectief. Echter, tegenwoordig zijn de onderwijskundige- en organisatorische kant net zo belangrijk. Dat wil zeggen, instellingen in het hoger onderwijs willen weten wat de effecten van learning analytics op het leren zijn en hoe learning analytics op grotere schaal kunnen worden geadopteerd. Om adoptie te ondersteunen zijn verschillende learning analytics modellen ontwikkeld. Deze helpen instellingen om learning analytics te gaan gebruiken. Toch hebben de bestaande modellen een aantal tekortkomingen. Zo zijn ze niet allemaal empirisch gevalideerd, missen zij soms praktische ondersteuning, is het onduidelijk hoe kritische dimensies in de praktijk moet worden gebracht en focust een aantal modellen voornamelijk op instellingen in Angelsaksische landen.

Dit promotieonderzoek vult de bestaande kennishiaten op en helpt instellingen in het hoger onderwijs om learning analytics te adopteren door een *resource-based* perspectief te nemen. In de *resource-based view* staan middelen (*resources*) en organisatorische vaardigheden (*capabilities*) centraal. Middelen hebben betrekking tot data, technologie, mensen, financiering et cetera. Vaardigheden beschrijven hoe deze middelen moeten interacteren om een organisatie te bevoordelen. In tegenstelling tot middelen zijn vaardigheden niet overdrachtelijk en moeten zij door de organisatie zelf worden ontwikkeld. Dit heeft tot gevolg dat onderwijsinstellingen die learning analytics willen adopteren zelf de vaardigheden moeten ontwikkelen die aansluiten op hun specifieke context. Zonder de juiste vaardigheden kunnen studentdata niet worden omgevormd in inzichten en verbeteringen aan het onderwijs. Dit roept de vraag op welke vaardigheden voor learning analytics van belang zijn en hoe deze in de praktijk moeten worden ontwikkeld.

Het promotieonderzoek geeft antwoord op de hoofdvraag: "*Welke aan informatiesystemen gerelateerde organisatorische vaardigheden voor learning analytics bevoordelen docenten en studenten binnen Nederlandse instellingen in het hoger onderwijs?*" Om antwoord te geven op deze vraag is een vaardighedenmodel voor learning analytics ontwikkeld door het toepassen van *Design Science Research* principes. Er is gebruik gemaakt van een *multimethod* benadering. Het proefschrift bestaat uit vier samenhangende delen. Het eerste deel beschrijft de motivatie voor het onderzoek en hoe de effecten van learning analytics op het leren gemeten kunnen worden. Het tweede deel demonstreert het belang van de juiste vaardigheden op het succesvol gebruik van learning analytics. In het derde deel wordt het Learning Analytics Capability Model ontwikkeld en aangescherpt. Het vierde deel relateert aan de ex-post evaluatie van het model en bediscussieerd de uitkomsten van het onderzoek.

Instellingen in het hoger onderwijs die learning analytics willen toepassen hebben veel uitdagingen. Een exploratieve *case study* toonde aan dat het ontbreken van vaardigheden met betrekking tot data, technologie en stakeholders serieuze gevolgen kan hebben bij de adoptie van een learning analytics instrument. Een andere *case study* benadrukt de noodzaak om gedegen te overwegen welke stappen gemaakt moeten worden tijdens het opschonen van data. Het opschonen van data kan effect hebben op de uitkomsten van analyses en daarmee op de daaropvolgende learning analytics interventies. Daarom moeten stakeholders geïnformeerd worden over de genomen stappen, waarom deze genomen zijn en wat de consequenties hiervan zijn. Wanneer learning analytics in de praktijk wordt toegepast rijst de vraag wat de effecten hiervan op het leren zijn. Immers, learning analytics hebben als doel het leren te verbeteren. Volgens leertheorie kunnen de gemeten effecten betrekking hebben op drie categorieën: de leeromgeving, het leerproces en de leeruitkomsten. Gebaseerd op 62 *key studies* die gevonden zijn tijdens een systematisch literatuuronderzoek zijn de categorieën uitgebreid met elf subcategorieën. De uitkomsten van het literatuuronderzoek helpen instellingen in het hoger onderwijs om de effectiviteit van learning analytics systematisch te beschrijven. Dit stimuleert vervolgens de adoptie van learning analytics door andere instellingen.

Onderzoek naar *resource-based* vaardigheden voor learning analytics is nieuw. Daarom is tijdens een systematisch literatuuronderzoek kennis uit meer volwassen onderzoeksgebieden zoals business analytics en big data analytics vertaald naar het learning analytics domein. Door het analyseren van tien *key studies* betreffende business analytics en big data analytics blijken vier categorieën van vaardigheden van belang: Data, Management, Mensen en Technologie. Echter, zes extra *key studies* uit het learning analytics domein tonen een vijfde categorie die niet moet worden vergeten: Privacy & Ethiek. Op basis van de uitkomsten van het literatuuronderzoek is een theoretisch Learning Analytics Capability Model met 34 vaardigheden ontwikkeld. Deze vaardigheden zijn geclassificeerd binnen de vijf eerdergenoemde categorieën. Om te helpen bij het in de praktijk brengen van het model bevat deze 461 operationele definities die gebruikers ondersteunen bij het omwikkelen van de vaardigheden binnen hun instelling. Hierna werd *case study* bij een instelling die volwassen is in het gebruik van studentdata uitgevoerd. Hierdoor kon empirische data aan het model worden toegevoegd. Het aantal operationele definities steeg naar 818. Ook werden zeven vaardigheden samengevoegd en drie hernoemd. Verder werden naar aanleiding van de *case study* de definities van alle vaardigheden verbeterd.

Als laatste stap in het onderzoek werd het Learning Analytics Capability Model door echte gebruikers geëvalueerd tijdens vijf *pluralistic walkthroughs*. In totaal hebben 26 praktijkbeoefenaars van vijf onderwijsinstellingen het model gebruikt om een implementatieplan dat zij voor hun instelling hadden gemaakt te verbeteren. Een aanvullende evaluatie werd uitgevoerd met zeven learning analytics experts en door het enquêteren van zowel de praktijkbeoefenaars als de experts. Het evaluatieonderzoek bevestigde de effectiviteit, bruikbaarheid en compleetheid van het model.

Het promotieonderzoek leidt tot verschillende, positieve uitkomsten. Vanuit praktisch perspectief biedt het Learning Analytics Capability Model een model voor Nederlandse instellingen in het hoger onderwijs en ondersteunt het stakeholders bij hun taak om de adoptie van learning analytics te plannen. Daarnaast voorziet het onderzoek in maatstaven om de impact van learning analytics op het leren vast te kunnen stellen, kunnen de voordelen van learning analytics beter worden bepaald. Dit helpt bij de volwassenwording van het learning analytics domein. Ook geeft het model duidelijke aanbevelingen met betrekking tot het opschonen van data. Niet alleen beroepsbeoefenaars maar ook academici profiteren van het onderzoek. De bestaande kennis met betrekking tot de vaardigheden belangrijk voor learning analytics wordt erdoor uitgebreid. Het Learning Analytics Capability Model probeert niet om bestaande modellen te vervangen. Het zet echter het goede werk voort en verbetert verschillende tekortkomingen die bestaande modellen hebben. Niet alleen combineert het Learning Analytics Capability Model wat al bekend is over learning analytics vaardigheden, het beschrijft ook hoe deze vaardigheden in de praktijk kunnen worden gebracht. Toekomstig onderzoek moet vaststellen in welke volgorde de vaardigheden moeten worden ontwikkeld en kan de effectgroottes van de verschillende vaardigheden op uitkomsten van learning analytics identificeren. Daarnaast laat dit onderzoek zien dat *Design Science Research* en de *resource-based view* waardevolle inzichten bieden aan het learning analytics domein en dat *pluralistic walkthroughs* kunnen dienen als methode voor ex-post evaluatie.

# Appendix 1: Capability definitions

**Appendix 1:** Capability definitions

| **Category** | **Capability** | **Definition** |
|---|---|---|
| Data | Data usage | This capability describes in what way data analysis is used to improve education. It contains data aggregation, different kinds of analysis, the goals of the analysis, and the interventions performed based on the outcomes of the analysis. |
| | Quality | This capability describes what the characteristics of 'high-quality data' are and in what way the quality of data is secured. |
| | Reporting | This capability describes in what way the outcomes of data analysis are presented to the various stakeholders and what the requirements of the presentation are. Reporting comprises all one-way flows of learning analytics outcomes. |
| | Sourcing & integration | This capability describes what data sources are used to get the data and in what way different sources are integrated. |
| | Feedback on analytics | This capability describes the need for feedback from users. |

**Appendix 1:** Capability definitions

| Category | Capability | Definition |
|---|---|---|
| | Identifying benefits | This capability describes in what way learning (process, environment, or student performance) is improved, i.e., with what goal is an intervention performed? |
| | Capability management | This capability describes in what way existing organizational capabilities are managed, reconfigured, and developed. This includes the growth of existing resources and capabilities into new ones. |
| | Culture & readiness | This capability describes characteristics of a data-driven culture and the readiness factors necessary for the adoption of learning analytics. |
| | Funding & investment | This capability describes what kind of funding and investment is necessary (time, people, money) and the role of executives to secure the proper funding and investment. |
| | External environment | This capability describes all influences from outside the organization. This includes the use of material and tools from external parties, the hiring of external personnel, requests and demands from external (governmental) bodies, and sharing materials, knowledge, and experiences with other (higher) educational institutions. |
| Management | Performance monitoring | This capability describes in what way the performance of analytical processes and applications are measured. |
| | Responsibility & accountability | This capability describes how the responsibility and accountability regarding learning analytics are managed. That is, decide on who is responsible and accountable for what task and the leadership structure should be. This capability can relate to the responsibility and accountability for activities and tasks at either stakeholder group level (e.g., the learning analytics team) or individual level (e.g., the data scientist from the learning analytics team). |
| | Strategy | This capability describes how to align learning analytics with the (long-term) organizational vision, strategy, and policy. |
| | Policies & CoP | This capability describes in what way to design and maintain learning analytics policies and codes-of-practice. |
| | Implementation Deployment & Application | This capability describes in what way to plan the use of learning analytics, i.e., what activities to deploy before the implementation of learning analytics (systems), and what factors to consider when implementing and deploying analytics. |
| | Evidence-based & theory-driven | This capability describes how to include evidence and theory in the design of learning analytics and the performance measurement of learning analytics interventions. |

**Appendix 1:** Capability definitions

| Category | Capability | Definition |
|---|---|---|
| People | Collaboration | This capability describes the active cooperation between parties - either within a group of stakeholders or between different groups of stakeholders, and either internal or external. This capability also includes the mechanisms via which collaboration is achieved. |
| | Combined skills and knowledge | This capability describes the combined skills and knowledge stakeholders need to have to perform learning analytics activities and act on them accordingly. |
| | Communication | This capability describes the flow of information between stakeholders (groups). This includes communication between users and the party delivering the learning analytics about the needs and possibilities ('demand and supply'), the communication mechanisms, and the types of information that are shared between different stakeholders (groups). |
| | Stakeholder Identification & Engagement | This capability describes the identification of different (types of) stakeholders who should be involved, and for what purposes stakeholders should be involved in the design, implementation, and use of learning analytics. |
| | Training | This capability describes what training stakeholders should get to perform learning analytics activities and act on the outcomes. |
| Technology | Automation | This capability describes the role and mechanisms of automation to perform learning analytics activities. |
| | Connectivity | This capability describes in what way systems can be connected |
| | Infrastructure | This capability describes the required IT infrastructure for learning analytics |
| | System characteristics | This capability describes both the quantitative and qualitative requirements for the systems used to perform learning analytics activities |

**Appendix 1:** Capability definitions

| Category | Capability | Definition |
|---|---|---|
| Privacy & Ethics | Ethics | This capability describes in what way ethical issues should be considered |
| | Human decision-making | This capability describes the role of humans in decision-making based on data. |
| | Security | This capability describes how to secure data and information |
| | Transparency | This capability describes being transparent about the used definitions, decisions made during the process and context in which learning analytics was used. |
| | Legal compliance | This capability describes how to comply with the law |

# Appendix 2: Operational definitions

### Data – Usage

- Aggregate data to a higher level
- Aggregate student id, degree program, and year. Each combination is one row in the data set
- Aggregate data in the right way
- Analyze data in near-real or real time that allows responses to unexpected educational events
- Analyze information in near-real or real time
- Analyze large amounts of educational data to understand the past and current state for specific target variables
- Analyze semi-structured and unstructured data (e.g., imaging data, the meaning and context of human language and voice) to improve education
- Analyze social media data to understand current trends from a large population
- Analyze whether demands for course credits are suitable or not
- Analyze effects of class scheduling and performance
- Analyze student success
- Answering questions related to degree programs, like when most students enroll or relations between learning behavior and enrollment
- Do course analyses and see how students flow through their education
- Examine undetected correlations, patterns, trends between specific variables of interest across regions or facilities
- Explore the causes of occurred educational events from relational databases
- Identify correlations and patterns from diverse data to gain new insights
- Identify from what schools students come from and adjust marketing activities accordingly
- Search educational databases for all data related to student characteristics and conditions
- Track learning events based on the rules that built on educational data sets
- Track student data based on the rules that built on educational data sets
- Analyze that when students are delayed, how realistic is it to expect they get back on track
- Cluster students based on common characteristics
- Compare “what if” scenarios
- Compare students to each other and plotting study paths
- Provide comparative interpretation of similar student cases over time
- Correlate student evaluation and certain elements of the program
- Detect fraud, abuse, waste, and errors in education

- Early identify students who might get delayed
- Early identify study problems
- Identify best moment of contact with students based on data
- Identify correlation
- Identify students who lag behind
- Identify students with academic potential
- Identify students who need extra follow up
- Identify delayed students
- Build automated algorithms where they can support decision making for improving student success
- Monitor student progress on a daily basis
- Support personalized learning strategies
- Use learning analytics in a supportive manner (support student and teacher daily work) rather than a business manner
- Allow to predict student behaviors
- Calculate chances of dropping-out
- Compare cross-referencing current and historical data and its outcomes to predict future trends
- Intervene early based on predictions
- Build models like a generalized additive model (GAM)
- Generate a set of predictions about the effectiveness of various intervention options for students based on unique characteristics.
- Make a study plan and predict its realizability
- Perform “what if” analysis using predictive modeling
- Predict chances of dropping-out based on 20 variables
- Predict costs and income
- Predict patterns of education in response to students needs
- Perform a regression analysis
- Be focused on providing learners with data and information to self-regulate their own learning
- Create awareness about other learning practices and social networks around them that they can use to make to meaningful connections
- Generate support for discussions and confirm gut-feeling
- Help learners keep track of the effects of their experiments on their learning in which “treatments” are changes they make to tactics used in learning
- Help learners understand their learning rather than pointing out what they are doing wrong
- Improve communication towards students, for example about possible barriers in the near future

- Inform students about study progress
- Make informed decisions about study paths
- Provide educators with formative feedback which helps them improve their practice
- Provide feedback to all stages of the learning process (design, dean, teacher, gov't officials, learner, etc.)
- Provide insights in study path
- Use learning analytics to evaluate policies
- Move from individual prediction to supporting policymaking processes
- Provide analyses and visualizations to help policymaking
- Provide near-real or real time information on educational operations and services within educational institutions and across educational systems
- Run broad multidisciplinary studies that extract important insights from large amounts of student data.
- Support organizational work groups
- Enable innovation in services or products based on novel use of (or insights from) data
- Use external sources (data) always facilitates understanding of market conditions.
- Use external sources (data) always facilitates understanding of students 'demands.
- Use internal sources for educational management always minimizes total educational costs.

**Data – Feedback on analytics**

- Develop new analyses based on feedback
- Let learning analytics be judged useful by learners
- Let feedback change the way results are presented.
- Ask users to inform the learning analytics team when they see strange things
- Provide opportunities for students to provide feedback on results of analytics.
- Seek feedback from primary users through various channels.
- Ask stakeholders for feedback

**Data – Quality**

- Clean data first
- Ensure data are good and sorted
- Decide the scope of the project and thus the range of data.
- Enter data correctly, otherwise this might lead to inaccurate records.
- Ensure the use of data with the right quality and data that are clean
- Select data that will be fed back to different stakeholders.
- Filter double records, for example students who do two courses and therefore are shown twice.

- Check whether records are not registered in the right academic year but according to the actual data of an event.
- Ensure there is no input error in all data.
- Ensure the use of the right kind of data
- Input all internal data sources without omissions.
- Input all external data sources without omissions.
- Identify missing data
- Check whether all data is read and not only the first couple of rows.
- Ensure data sets are correct as problems due to incomplete data are hard to be found.
- Ensure that data sets are complete as this makes qualitative good data.
- Ensure there is no missing data.
- Ensure that the learning analytics used provides a complete set of information.
- Ensure that the learning analytics used provides all the information needed.
- Have data sets that are complete, so no other scenario's than the one showed are possible
- Ensure that the learning analytics used produces comprehensive information.
- Ensure that the learning analytics used produces correct information.
- Ensure that the learning analytics used provides accurate information.
- Ensure that the learning analytics used provides few errors in the information.
- Have standards for comparison across instances, institutions, etc.
- Use a common definition of the main data source.
- Manage all data in the same way throughout the organization.
- Standardize data to support comparisons across areas.
- Consider data needs, access and availability
- Provide easy access and use of data
- Make students records consistent, visible and easily accessible for further analysis.
- Ensure that the information provided by the learning analytics is clearly presented on the screen.
- Ensure that the information provided by the learning analytics is well formatted.
- Ensure that the information provided by the learning analytics is well laid out.
- Make clear what data are about.
- Standardize labeling, so you know what the data is about
- Add time stamps as most data sets don't have them.
- Prevent the use of open text fields as they lead to ambiguous labeling of data.
- Ensure the meaning of data labels are clear.
- Ensure the labeling of variables is unambiguous.
- Ensure all variables are described in documentation.
- Ensure data can be linked, otherwise the usability is minimal at best

- Construct linking tables to link data from different sources
- Extract data in such way that is easily maintainable
- Have master data management and metadata management to ensure data definitions
- Have mechanisms to ensure quality
- Compare input and output automatically
- Ensure that the learning analytics used always provides up-to-date information.
- Ensure that the learning analytics used produces the most current information.
- Ensure that the learning analytics used provides the most recent information.
- Realize that when data are inputted not on time, the analysis cannot lead to valid outcomes

### Data - Reporting

- Combine data to get new insights
- Generate proactive learning recommendations
- Provide actionable insights or recommendations in a format readily understood by its users
- Provide actionable insights to decision-makers in near-real or real time
- Provide system outputs for role-based decision-making
- Ensure that reports are in the right format and show the right data to inform decisions
- Provide actionable actions as there is a gap between getting new insights and knowing what to do next
- Build a good dashboard that provides insights
- Build a good dashboard as a lot of data is stored in Excel sheets but a dashboard with filters works quicker and looks nicer
- Build a good dashboard
- Use, for example, Tableau
- Build a good dashboard or spreadsheet with filters
- Build a good dashboard to show the study advise
- Generate learning summaries (or performance metrics) in real time or near real time and present this in visual dashboards/systems
- Make a dashboard to see where in a program students earn their course credits
- Realize that with the more insights, the more variables are necessary, the more complex the dashboard gets
- Choose analytical models and define metrics.
- Make clear how confident any predictions about student success/failure are, and why those predictions are being made
- Justify models with text that is derived from the dashboard
- Explained outcomes so no wrong conclusions are drawn

- Report also the programming codes and scripts
- Describe the justification also in text when reporting a model.
- Give a demonstration
- Give internal presentations
- Present results
- Think about the way outcomes are presented.
- Limit the number of variables included in a report in order to come closer to the user.
- Realize that it is hard to exactly present what the data mean and that presentation is always a bit clouded.
- Produce results that are presented in such a way that the end users can drill down to the raw data so that validity can be checked.
- Provide systemic and comprehensive reporting to help recognize feasible opportunities for learning improvement.
- Provide a visualization with key figures
- Consider the best ways to present analytics results (e.g., visualization).
- Visualize learning analytics outcomes in pictures
- Provide visualizations
- Use visualizations with colors
- Provided tutors and student advisors with an overview

**Data – Sourcing & Integration**

- Build the ability to acquire, curate, maintain and enhance datasets based on value maximization
- Have access to historic learning results
- Have access to very large, unstructured, or fast-moving data for analysis
- Be able to capture data from all sources and collect this data in a single repository
- Go to the source when data are not in the central database
- Link data warehouse to dynamic, learning management system data, and other forms of student experience across the campus
- Build mechanism for integration of existing data in a central repository, e.g., a data warehouse.
- Source data from the Contact Management System
- Source data from the Management Information System
- Source data from the Virtual Learning Environment
- Source data from the Enterprise Resource Planning (ERP) system
- Source the Contact Management System for, e.g., notes from study advisors.
- Source not only the Virtual Learning Environment but also systems of the administration department
- Source, for example, notes from chats with students.

- Collect data from the Management Information System and the Enterprise Resource Planning (ERP) system
- Extract data from the Management Information System
- Use administrative data
- Program the linking of data in, for example, R.
- Manually repair errors in linking tables.
- Realize that there are a lot of customizations that need to be done to the data from the original source.
- Take care that, when combining data sets, records are not doubled because, for example, students are counted twice.
- Be careful for correlation between imported variables.
- Realize that when linking tables are not updated, for example when the name of a degree program changes and this is not implemented in all tables, errors occur.
- Collect data from external educational sources and from various learning systems throughout your organization
- Continuously share and cooperate using external sources (data) in order to make predictions.
- Integrate data from other educational institutions and data sources.
- Integrate seamlessly educational data across multiple regions or facilities in near real time or real time.
- Integrate external data with internal to facilitate high-value analysis of our institutional environment.
- Integrate heterogeneous data from multiple learning systems and devices
- Build mechanism for sourcing data for learning analytics initiatives from multiple channels, including operational/transactional systems and third-party sources.
- Integrate learning analytics systems with operational/transactional systems at the process, technology, and data levels in order to exploit the capabilities of both.
- Combine different types of data from all areas of the educational institution.
- Integrates data from different areas of the educational institution.
- Pull together data that comes from different places in the educational institution
- Integrate data from multiple internal sources into a data warehouse or mart for easy access
- Use numbers (ID numbers) to link data sets when all personal data are removed
- Make linking tables
- Link afterwards, like gender. These data come from other sources but can be included.
- To identify students, different codes might need to be used.
- Link data
- Use linking tables

**Management – Capability Development**

- Aggregate data to the right level
- Let learning analysts have sessions together so they can better align how to report on results
- Build capacity to analyze and predict data value trends or sources from current indicators, patterns and context
- Decide with the learning analytics team how to report on results
- Let learning analysis calculate whether results are significant, as users often don't ask this themselves.
- Develop ways of communication.
- Build the ability to reconfigure and leverage an organization's learning analytics resources and capabilities in order to respond to changes in the business environment in a timely and efficient manner. Such responsiveness requires the ability to identify potential learning analytics opportunities (Search), prioritize those opportunities based on business need, risk and technology maturity (Select) and then funding and implementing the opportunities (Asset Orchestration) resulting in new and unique resource configurations
- Build analytics capability
- Level the integration of data value into IT leadership and governance infrastructures
- Mature the learning analytics team
- Incorporate learning analytics into educational practices
- Revise the policy when learning analytics is used for new goals
- Define the role of learning analytics team
- Build the ability to evolve, adapt and orchestrate the data processes, support services and workflows towards value-based goals and outcomes
- The ability to enhance organizational management and technical skills about data value issues
- Evolve the organization based on data value considerations
- Provide lots of room for experimentation
- Transform the learning analytics activities
- Have confidence that learning analytics project proposals are properly appraised.
- Appoint specialists to lead learning analytics projects
- Identify required expertise (e.g., learning analytics expertise, IT expertise, statistical expertise, educational expertise, psychological expertise)
- Share knowledge now possessed by a single person, e.g., via protocols
- Be careful with advice as the learning analytics team doesn't know the contextual causality
- Let the learning analytics team decide for itself what competencies must be further developed

- Have sufficient contact with users
- Transfer knowledge to the organization, for example via a business game
- Limit the level of detail of an analysis to come more towards the users' need
- We want for every stakeholder group one person who truly understand it
- Limit the damage when a dataset is leaked by removing personal data
- Facilitate learning analytics through an enterprise-wide system rather than software-centric solutions
- Start with an institutional capability assessment encompassing dimensions such as institutional risk appetite, faculty perceptions, student consultation, data quality audit and technical capability
- Develop the ability to secure organizational mandate
- Bring together traditionally disparate organizational units within an institution in a project context that is fluid, rapidly changing, and relatively unchartered
- Build learning analytics capacity
- Do calculations centrally rather than locally, for example on a R server
- Use collaboration software like Slack to share knowledge

**Management – Culture & Readiness**

- Ensure that the administration largely accepts the use of analytics
- Ensure the faculty largely accept the use of analytics
- Have a 'learning analytics ambassador'.
- Support willingness to embed these processes and practices in daily work
- Change how key stakeholders perceive the project
- Change how something is done: how decisions are made, how learners are supported
- Demonstrate culture change through changed behavior.
- Make sustainable changes in the way student success is achieved or supported
- Ensure that learning analytics is recognized as crucial to research in teaching and learning
- Ensure that learning analytics is recognized as valuable by those who do teaching (instructors, instructional designers, curriculum committees)
- Let more and more people find learning analytics useful
- Show there is a latent need: people don't ask for it but they see it usefulness
- Convince people about the usefulness of learning analytics
- Get new ideas how to better support students, thereby adding value compared to the old situation.
- Suppress the feeling that when things go well, there is no urgency for change
- Have learning analytics be very widely implemented in an ecosystem that shares data
- Create a culture where formal authority, reputation, intuition, and ad-hoc decision-making are preceded by decisions based on data

- Create a culture where learning analytics users, including power users, are encouraged to actively participate in the development of a data-driven environment
- Create a culture where whenever possible, assertions are substantiated with data
- Create a culture where although the emphasis is on fact-based decision making, there is still some room for intuition and ad-hoc decision-making, particularly when the required data is not available
- Create a data-driven mind-set
- Incorporate a data-driven mind-set in processes
- Create the feeling that people can override their own intuition when data contradict their viewpoints
- Create a feeling that people can base their decisions on data rather than on instinct
- Let people consider data a tangible asset
- Coach employees to make decisions based on data
- Create a culture that accepts the use of data to make decisions; we are not reliant on anecdote, precedent, or intuition
- Create a system that can generate arguments to support your feeling.
- Let learning analytics be accepted as a standard of practice
- Use formative feedback-based rewards for staff (rather than summative performance indicators with penalties)
- Discourage teaching staff and students from gaming the system.
- Enable experimentation to discover needs
- Use learning analytics consistent with institutional practices
- Continuously assess and improve the business rules in response to insights extracted from data
- Have a process for moving from what the data say to making changes and decisions
- Create the ability to manage human, technological and process impacts across the organization arising from learning analytics initiatives. This involves the provision of training and rewards in order to demonstrate the value and utility of learning analytics
- Create the ability to manage human, technological and process impacts across the organization arising from learning analytics initiatives. This involves the provision of training and rewards in order to manage expectations
- Make everybody aware of the possibilities and start to expect learning analytics
- Create the ability to manage human, technological and process impacts across the organization arising from learning analytics initiatives. This involves the provision of training and rewards in order to encourage the adoption of new learning analytics technologies and work practices
- Measure the extent to which learning analytics permeated the fabric of an organization e.g., learning processes and values. Where appropriate, models are used to make decisions on an ongoing and pervasive basis

- Measure the extent to which learning has permeated the fabric of an organization e.g., learning processes and values (e.g., appreciation for learning analysis tools and data-driven insights). It is reflected in the extent to which people routinely use data and learning analytics tools to solve problems and make decisions.
- Realize that learning analytics capabilities are not transferable
- Embed processes and practices in the fabric of the institution and use them effectively.
- Use learning analytics to fit the organizational culture
- Change from thinking in terms of money and efficiency towards thinking in quality like student welfare and whether students learn what the job market expects from them
- Offer rewards for use to students and teachers that overcome inherent inertia and resistance
- Create the ability to manage human, technological and process impacts across the organization arising from learning analytics initiatives. This involves the provision of training and rewards in order to mitigate potential resistance
- Realize there might be resistance as learning analytics is applied to an unknown context
- Have complete support and recognition from the institution's senior leadership
- Develop strategies to engage leaders, promote buy-in and change educational culture
- Embrace the whole system with guidance from key leadership
- Raise awareness and understanding of learning analytics among teaching staff and students through publicity and meetings/ workshops/ conferences.
- Let top-management promote the use of learning analytics in the organization.
- Have senior leaders who are interested in and committed to using data in making decisions
- Establish a culture where there is trust in data and the learning analytics tools used to analyze data
- Evaluate institutional culture (e.g., trust in data and openness to changes and innovation)
- Let learning analytics personnel across the organization foster a culture of open communication and trust between themselves and other users.
- Beware of events that lowers the trust in learning analytics
- Research what are reasons for people to distrust the data
- Get people to trust learning analytics

**Management – Evidence-based & Theory-driven**
- Advance pedagogies that educators value
- Let learning analytics be driven by pedagogy

- Identify reasons for learning improvement
- Formulate questions for specific improvement, otherwise you get lost in the data
- Blend learning analytics with proven best practice
- Motivate faculty to make a change in their pedagogy because the evidence of impact on student learning is clear
- Engage with existing learning analytics cases and literature.
- Realize that the more nuanced the analyses are, the less evidence-based their outcome is
- Don't be too nuances, otherwise it becomes less interesting for final users
- Use existing evidence that could convince others of the need for change.
- Develop a theoretical and conceptual foundation that unifies research from many disciplines interested in learning and teaching
- Support learning analytics with (pedagogical) theory

**Management – External Environment**

- Collaborate with other institutions
- Consult external expertise
- Use relevant materials of a third party to formulate policy
- Visit relevant conferences
- Compare with our rivals within the educational industry whether our organization has the foremost available analytics systems.
- Research to what extent the 'suppliers' of your current students (secondary schools, higher educational institutions) have implemented learning analytics
- Research to what extent competitors have implemented learning analytics
- Research to what extent partners have implemented learning analytics
- Understand new market s
- Research whether there are influences or demands from government
- Identify students from other institutions, from who previous data is not known

**Management – Funding & Investment**

- Let funding for analytics be viewed as an investment in future outcomes rather than an incremental expense
- Consider and estimate the time managers will need to spend overseeing the change.
- Consider and project about how much these options will help end-users make quicker decisions.
- Think about and estimate the cost of training that end-users will need.
- Think about and estimate the effect they will have on the productivity of the employees' work.

- Have executive commitment to investing in new tools, solutions, and practices and especially in changing the culture and behaviors
- Let financial resources be allocated by directors
- Get management buy-in
- Develop a human and fiscal resource investment plan, including a long-term commitment to launching, resourcing, scaling, and sustaining the effort
- Account for limitations of time, resources, and/or budget
- Have central investment
- Create the ability of senior managers and executives to advocate the use of learning analytics systems and data-driven decision-making throughout the organization. This requires (iv) the provision of financial and material support for learning analytics initiatives
- Be aware for a potential decrease in resources due to the inconvenience of sharing educational information about internal sources (data)
- Evaluate financial capacity
- Secure funding to hire personnel
- Invest in additional personnel
- Secure capital/financial resources
- Secure adequate funding for learning analytics projects
- Give learning analytics projects enough time to achieve their objectives
- Make resources available to substantiate learning analytics
- Seek funding
- Avoid exceeding budgets
- Be aware for time constrains
- Seek an appropriate amount of funding for analytics
- Hire an appropriate number of analysts for analytics

### Management – Identifying Benefits

- Identify the changes in organizational goals based on data value
- Define targets at course level, not more detailed than that
- Identify the key outcomes the organization is trying to improve with better use of data
- Address the problems and requirements experienced during learning and teaching, rather than those of institutions and researchers
- Demonstrate empirical impact on student success
- Provide evidence of high improvements in learning experiences
- Improve customer service/student care
- Cluster populations to customize actions
- Evaluate the program based on information

- Have a chat with students with high risk of dropping out
- Act early on problems later in the curriculum
- Improve the curricula
- Minimize workload through simple visualizations available to both faculty and students
- Expose variability and improve performance
- Improve the quality of work
- Improve planning
- Make work more efficient
- Demonstrate value (ROI) to institutions that do not have retention or performance issues
- Use internal resources (data) to decrease management costs and time for planning and education
- Lower costs
- Share information about internal sources (data) to reduces educational costs.
- Grow sales to new customers or new markets
- Identify new product/service opportunities
- Innovate through new business models, products, and services

**Management - Implementation, Deployment & Application**

- Deploy learning analytics in a systemic manner with the needs of varying stakeholders accounted for
- Consider contextual elements (e.g., institutional size, structure) to identify opportunities for learning analytics
- Develop a plan that prioritizes selected stakeholders and purposes
- Identify areas where different stakeholders will be supported by learning analytics (macro level –institution, meso level –department/ programme, and micro level – teaching staff and students)
- Identify expected 'changes' to the current context and key stakeholders (e.g., teaching staff and students)
- Make learning analytics specific for the different degree programs
- Let the learning analytics team have a serving purpose
- Continuously evaluate and make adjustments to the work of the learning analytics team
- Have a short adaption cycle, so nobody tends to do things themselves
- Spend time to reflect on what users really want
- Implement learning analytics bottom up
- Enable implementation that recognizes the complexity of educational systems

- Recognize the need for cross-institutional understanding, skills upgrades, and acceptance when implementing learning analytics
- Realize that at first, the technical and privacy aspects make it hard to start learning analytics project
- Consider phases of implementation (e.g., explore data, carry out pilot projects, seek feedback from users, and develop a policy for the adoption of learning analytics).
- Evaluate risks
- Mind inadvertent consequences and make sure the benefits of learning analytics to students outweigh risks
- Plan a learning design strategy to maximize production of meaningful data
- Make learning analytics projects practical and implementable, e.g., not just narrow research but lending itself to system/program design and delivery
- Make a preliminary action plan and timelines which can be reviewed and modified as needed.
- Create top-management support for learning analytics initiatives within your organization?
- Enforce adequate plans for the introduction and utilization of learning analytics.
- Frequently adjust learning analytics plans to adapt to changing conditions
- Have a list of things to do for the institution
- Make a planning, so everybody can see when which theme is worked on
- Make a planning per year, including organizational goals
- Perform the learning analytics planning processes in systematic and formalized ways.
- Use version control, so you can go back in case a mistake is made
- Write a project proposal
- Integrate learning analytics in daily learning events
- Integrate learning analytics into the environments, practices and processes of teachers and students
- Integrate learning analytics with the work practices of educators
- Reflect the complexity and multi-dimensionality of teaching and learning practices
- Consider whether there is political interest in change, and how key decision makers may perceive the problem(s) that learning analytics could address and/or the proposed solutions
- Embedded learning analytics in existing functions
- Engage with research projects locally or through collaboration with other institutions
- Identify opportunities to build learning analytics upon existing projects or practice.
- Have a value-centric data process design and integration of data value considerations into wider organizational processes
- Decide on forms of interventions (e.g., automatic systems, personal contacts, learning resources)

- Plan strategies to intervene at points identified in the context-mapping work

**Management – Performance Monitoring**

- Perform a policy, processes, and practices audit
- Develop methods to triangulate analytics results
- Build organizational capacity to create a data value-based control loop that feeds into organizational level decisions
- Not just encourage but orchestrate and measure student success, with a focus on continuously improving results
- Support data value assessment and reporting at all stages in the data value chain
- Constantly monitor the performance of the learning analytics function.
- Establish indicators of data quality and system efficacy
- Establish qualitative and quantitative indicators of success
- Be clear about the learning analytics performance criteria.
- Set up measurable milestones

**Management – Policies & Code-of-Practice**

- Consult relevant policies and codes of practice (e.g., Jisc's Code of Practice for Learning Analytics, and data protection policies)
- Describe the aims of learning analytics at the institution in a Code of Practice
- Have a Code of Practice to make clear what is done with learning analytic
- Check whether the original policy objectives and vision are still accurate and relevant in the light of the assessment of context, purposes, and capacity
- Change written policy with regard to evidence-based support of learners

**Management - Responsibility & accountability**

- Create independent learning analytics units
- Enhance learning analytics through the provision of some authoritative autonomy and financial independence, which provides learning analytics managers with a degree of freedom to pursue value-creating actions.
- Split learning analytics across multiple organizational units
- Control learning analytics projects centralized
- Control learning analytics projects decentralized
- Establish a diverse working group (including teaching staff and students) and define a clear leadership structure
- Have complete new learning analytics activities approved by the board, with the freedom to make minor changes to the current concept.
- Be aware of whom is deciding where money is spend on, for example, the educational director

- Define a role for the learning analytics team: are they internal consultants or just the data suppliers?
- Assign decision rights and responsibilities by determining; those who will be held accountable for the resulting actions and outcomes of these decisions. It is important that a person responsible for making a certain decision is held accountable for the resulting actions and outcomes
- Assign decision rights and responsibilities by determining; those responsible for making certain decisions in relation to the planning, implementation and applications of learning analytics
- Assign decision rights and responsibilities by determining; where appropriate, those who will provide the input for such decisions.
- Let the learning analytics team agree on the role they have
- Specify who is responsible for learning analytics development
- Ensure every member of the learning analytics team can do multiple jobs, as roles within the team can change.
- Manage access rights, for example via the IT department
- Let the privacy officer make sure everything that is done is allowed.
- Let the security officer check whether everything the learning analytics team does is done in a secure way.
- Made a planning, so everybody can see when a specific theme is worked on.
- Specify what the position of the learning analytics team within the organization is; what can they do and what not
- Don't have the learning analytics team act too much as consultant as they are not the substantive expert
- Describe what is inside an algorithm, as this is demanded by privacy regulation
- Let have the learning analytics team look into whether a difference is significant or not, as policymakers often do not ask this themselves.
- Carefully consider the advice the learning analytics team gives and also make sure people don't interpret the analyses too firmly
- Promise that data is collected in a reliable and reproducible way, that the analyses are valid, and that analyses are delivered in the form of a report.
- Don't give mere advice based on what is discovered but point out interesting insights and ask users to think about that themselves.
- Let the academic advisors contact students
- Let the focal department act upon the insights gained from learning analytics
- Let the organization act upon insights given by the learning analytics team.
- Have tutors and academic advisors look further than the dashboard and ask deeper questions.

- Don't have the learning analytics team involved in the operational part, that is, which students should be contacted now.

**Management – Strategy**

- Align with organizational design
- Let learning analytics be integrated into the long-term planning that drives the institution
- Meet identified institutional needs
- Transform organizational design
- Align the organization's learning analytics initiatives with its business strategy. It is a two-way relationship in the sense that learning analytics initiatives can help measure and enforce a business strategy, whilst business strategy necessarily shapes learning analytics initiatives as they evolve. This requires a clearly defined business strategy that is enunciated to all staff and translated into a set of measurable outcomes. It also requires a genuine commitment to the strategy demonstrated by the decisions and actions of senior people
- Identify the university's learning and teaching strategies
- Let the long-term institutional goals include the implementation of analytics strategies, tools, and processes to serve multiple stakeholders
- Prioritize stakeholders and goals realistically in order to meet institutional goals,
- Continuously examine the innovative opportunities for the strategic use of learning analytics.
- Make a planning for a year, including organizational goals
- Align learning analytics with the wider institutional strategies or introduce learning analytics into the university's strategy
- Identify important business insights and trends to improve educational services
- Identify internal and external drivers for learning analytics (e.g., problems to solve or areas to enhance).
- Let learning analytics help developing organizational policy
- To what extent has the top-management promoted learning analytics as a strategic priority within your organization?
- Create the ability of senior managers and executives to advocate the use of learning analytics systems and data-driven decision-making throughout the organization. This requires a clear vision
- Create the ability of senior managers and executives to advocate the use of learning analytics systems and data-driven decision-making throughout the organization. This requires the promotion of this vision and understanding throughout the organization
- Identification of longer-term data acquisition, data architecture, system integration, IS priorities, business process and people required

- Create the ability to specify sources of value for a specific organization
- Have program directors on board to ease the change of program-specific policies

**People – Collaboration**

- Let individuals with access to policy makers use their connection to reach out to important stakeholders.
- Let an expert, e.g., a mathematics professor, checks the used methods
- Invite other institutions to do an 'internship' with to learn about what done with regards to learning analytics or vice versa
- Have different departments and levels of the organization come together to discuss matters
- Involve different scientists
- Get support from experts like a mathematics professor
- Have intake interviews to make sure the learning analytics team on the same level as the principle of a learning analytics related project
- Inform principles of learning analytics related projects when they request for learning analytics will be handled and what we are going to do.
- Involve scientists when conducting experiments, so they can publish papers about it.
- Have the dean and even the board reading reports first before publishing.
- Have contact with principles of learning analytics related projects all the time, otherwise things go wrong
- Let the learning analytics team have frequently short meetings to share what they are working on
- Let team members of the learning analytics team identify what they find important to write in learning analytics reports
- Make "openness" a priority so that models, data, and best practices can be shared across institutions
- Extent to which learning analytics has permeated the fabric of an organization e.g., business processes and values. It is facilitated by sharing metadata and the use of a collaboration portal.
- Realize that for users alone, the whole learning analytics process is too complex to fully understand
- Proposed to other users to help them and also collaborate with the learning analytic team
- Let learning analysts and line people coordinate their efforts harmoniously.
- Involve all types of learning analytics users, from managers to operational staff, in the initial planning of a learning analytics initiative
- Have analytics personnel work closely with users and maintain productive user/client relationships.

- Let the learning analytics team work together with the users to show and explain them what is done
- Let the learning analytics team think along with the users about their needs
- Continuously communicate between users and learning analytics team
- Work together with the users
- Enable users to ask questions to the learning analytics team.

**People - Combined Skills & Knowledge**

- Let academic advisors have their own way of identifying and contacting students
- Interpret it in the right way
- Realize that not everybody fully understands the outcomes, as many people are used to make decisions on a table from the management information system and their own interpretations
- Have stakeholders understand how the system works, otherwise they might just copy it to Excel so they can do their own analyses.
- Realize that the users do not always find the data and dashboard easy to understand and to use
- Provide users with knowledge about the dashboard so they can do more with is
- Develop combined skills and knowledge of learning analytics managers and other learning analytics users throughout the organization to (iii) create and promote a technical innovation team
- Develop combined skills and knowledge of learning analytics managers and other learning analytics users throughout the organization to (iv) an innovation forum made up of innovation teams from other business units.
- Evaluate human capacity (e.g., data literacy, relevant expertise, staff workload, opportunities for skill transfer)
- Have a mix of hands-on people, people who knows how to write articles, and people with business intelligence experience
- Have teams comprise specialists whose skills are complementary to other team members
- Have learning analytics managers who are able to coordinate learning analytics-related activities in ways that support other functional managers, suppliers, and students
- Have learning analytics managers who are able to work with functional managers, suppliers, and customers to determine opportunities that learning analytics might bring to our institution
- Let the data scientist do the analyses and have the business analyst writing the story around it
- Write a protocol to prevent only one person having the specific knowledge

- Be able to acquire new and relevant knowledge
- Be able to apply relevant knowledge
- Be able to assimilate relevant knowledge
- Be able to search for new and relevant knowledge
- Make concerted efforts for the exploitation of existing competencies and exploration of new knowledge.
- Develop combined skills and knowledge of people in learning analytics related management roles throughout the organization to prioritize and manage learning analytics projects
- Have learning analytics personnel who are very capable in terms of managing project life cycles.
- Have learning analytics personnel who are very capable in terms of planning and executing work in a collective environment.
- Have learning analytics personnel who are very capable in terms of planning, organizing, and leading projects.
- Perform project evaluation
- Develop combined skills and knowledge of people throughout the organization that are involved in the business side of learning analytics initiatives, that also includes the ability to network
- Develop combined skills and knowledge of people in learning analytics related management roles throughout the organization to translate, communicate and sell the potential values and benefits of learning analytics to senior executives (e.g., senior executives and general managers).
- Have institutional research professionals who know how to support analytics
- We have IT professionals who know how to support analytics
- Have skilled resources
- Have learning analytics staff with the right skills to accomplish their jobs successfully
- Have learning analytics staff that holds suitable work experience to accomplish their jobs successfully
- Have dedicated professionals who have specialized analytics training
- Hire new employees that already have learning analytics skills
- Create the ability of senior managers and executives to advocate the use of learning analytics systems and data-driven decision-making throughout the organization. This requires first-hand experience and understanding of the benefits and successes of learning analytics
- Have evidence of knowledge of the field of learning analytics
- Have one person for each group of stakeholders who truly understands it all.
- Have analytics personnel that is very knowledgeable about the role of learning analytics as a means, not an end

- Have learning analytics managers that are able to understand and evaluate the output extracted from learning analytics
- Realize that a lot of people don't know the difference between anonymizing and pseudonymising data
- Develop combined skills and knowledge of learning analytics technology specialists across the organization including; programming, optimization software, algorithms, database/file management, ETL (Extraction, Transformation and Loading), data warehousing, software development methodologies and high level architectures. Some level of business domain and industry knowledge is necessary to apply these skill sets.
- Have learning analytics staff that can do data manipulation
- Have learning analytics staff that that can program
- Have learning analytics personnel that is very capable in terms of programming skills.
- Have learning analytics personnel that is very capable in the areas of data and network management and maintenance.
- Have learning analytics staff that can program, for example in R
- Realize that a lot of people don't know the difference between anonymizing and pseudonymising data
- Let stakeholders have understanding of analytics design and development
- Let stakeholders have understanding of data science
- Even when the data are visualized in a good way, ensure that people know what they are looking at
- Let stakeholders have understanding of qualitative analysis
- Increase statistical knowledge
- Increase knowledge about text mining
- Have the right statistical knowledge
- Ensure that learning analytics staff is able to interpret a boxplot
- Learning analytics team members need to know what they are doing scientifically
- Ensure that learning analytics staff are very knowledgeable about educational functions.
- Ensure that learning analytics staff are very knowledgeable about the learning environment.
- Ensure that learning analytics managers have a good sense of where to apply learning analytics
- Ensure that learning analytics staff are very capable in terms of teaching others
- Collaborate educational experts, as they know a lot about higher education and often have good ideas
- Collaborate with experts on how to effectively write an articles
- Let have learning analytic team members understand data visualization

- Let have learning analytic team members understand institutional reporting and/ or business intelligence
- Let have learning analytic team members understand learning analytics reporting
- Ensure that analytic tools are only used by people with sufficient understanding of how they work so that the results can be treated with an appropriate level of skepticism
- Let learning analytic team members have ability to use automated support processes for student success
- Let have learning analytic team members understand database development
- Let have learning analytic team members understand front-end / interface development
- Let have learning analytic team members understand IT support
- Let have learning analytic team members understand learning technologies administration
- Ensure that learning analytics staff possess superior understanding of technological trends.
- Ensure that learning analytics staff possess superior ability to learn new technologies.
- Ensure that learning analytics staff have the sufficient technical expertise to understand the data available to them and communicate with learning analytics technical specialists
- Ensure that people throughout the organization, that are involved in the business side of learning analytics initiatives, also possess the ability to seek out opportunities and threats.
- Ensure that people throughout the organization, that are involved in the business side of learning analytics initiatives, also possess the ability to develop and drive an agenda.
- Encourage learning analytics managers and other learning analytics users throughout the organization to continually challenge the status quo
- Ensure that learning analytics managers and other learning analytics users, throughout the organization, manage innovation as a separate activity to continuous improvement
- Let learning analytics related management roles throughout the organization redesign business processes as a result of implementing learning analytics
- Ensure that people throughout the organization, that are involved in the business side of learning analytics initiatives, know the depth of domain knowledge of the organization's key products, services, processes, value chain and industry in general.
- Ensure that people throughout the organization, that are involved in the business side of learning analytics initiatives, know fundamental business principles,
- Let have stakeholders understand data governance development and management

- Let have stakeholders understand data policy development and implementation
- Let have stakeholders understand engagement/management
- Let have stakeholders build an entrepreneurial mindset and vision and the ability to rationally assess risks and benefits.
- Ensure that analytics personnel are very knowledgeable about the critical factors for the success of our organization.
- Ensure that analytics personnel understand our organization's policies and plans at a very high level.
- Ensure that learning analytics managers are able to anticipate the future business needs of functional managers, suppliers, and students
- Ensure that learning analytics managers understand and appreciate the needs of other functional managers, suppliers, and customers.
- Ensure that learning analytics personnel are very capable in interpreting business problems and developing appropriate technical solutions
- Be careful with providing advice if the causality is not clear.
- Have learning professionals who know how to apply analytics to their area
- Ensure that the context the data is coming from is identified
- Ensure that you include contextual knowledge before drawing conclusion

**People – Communication**

- Signal remarkable things, so the users can think for themselves why this is.
- Don't just give insights and let people do the rest of the work themselves but collaborate to make the best from it
- Discuss how things work
- Discuss with people who work in that specific context how things work
- Create a culture of open communication and trust that involves listening carefully to the needs of users and translating learning analytics concepts into every-day language. It is facilitated by close and frequent contact via a variety of different communication channels
- Understand each other so things can be done better
- Conduct an intake interview with the principle to make sure we are aligned before beginning a project
- Discuss with stakeholders how they see the output of analysis
- Be able to offer different things depending on who the learning analytics team
- Ask users to inform the learning analytics team when they see strange things
- Keep communicating with users whether they want things different in (e.g.) dashboards that are designed

- Keep in touch with principles of learning analytics related project and validate their questions to make sure the learning analytics team researches exactly what they want.
- Let comprehensibility for users not come at the expense of detail in analyses.
- Let the learning analytics team write manuals to increase clarity in communication
- Ensure that users understand why we make a dashboard
- Give a lot of internal workshops
- Use different graphs for different degree programs
- Justify models in the texts
- Change how information is communicated and shared
- Establish communication channels between different stakeholders across the institution.
- Test your findings statistically before they are shared with stakeholders
- Let information be widely shared between the learning analytics team and line people so that those who make decisions or perform jobs have access to all available know-how.
- Let the learning analytics team and line people from various departments, frequently attend cross-functional meetings.
- Let the learning analytics team and line people meet frequently to discuss important issues both formally and informally.
- Identify communications bottlenecks within our organization when sharing analytics insights.
- Enable users to ask the learning analytics team to build a specific dashboard
- Have more and more people come with specific questions
- Enable users to ask the learning analytics team for specific information
- Be able to provide clear answers to users who come with specific questions.

**People - Stakeholder Identification & Engagement**

- Let learning analytics be co-designed with educators who understand what good learning looks like in their field
- Engage all stakeholders involved in teaching, learning, and governance of educational systems
- Involve students in the design and validation of new tools, possibly with pedagogical benefits to the students
- Set up a feedback session with students, tutors, academic advisors and other interested people
- Ensure broad stakeholder engagement
- Consider responsibilities and implications for all stakeholders
- Inform strategic planning and the design of approaches to involve, inform, support

and train key players
- Invite teaching staff to contribute their professional knowledge to the design and implementation of learning analytics (e.g., guide students to reflect on possible ways to act on the results of analytics)
- Ensure continuous communication between users and learning analysts
- Keep contact with faculties within the institution
- Identify academic advisors who can use learning analytics
- Identify academic advisors, policymakers, and communication staff who can use learning analytics
- Identify learning analytics ambassador
- Identify faculty directors who can support the use of learning analytics
- Identify academic teams (e.g., Learning & Teaching committee, Digital Learning Committee, research project teams)
- Identify external partners (e.g., researchers and service providers)
- Identify internal advocates of learning analytics among members of faculties (bottom-up approach).
- Identify primary users of learning analytics (e.g., students, teaching staff, and senior managers).
- Identify professional teams (e.g., IT, legal team, strategy team, Student Support, Student Registry, library).
- Identify senior management team (e.g., vice-chancellors, principals, provosts).
- Identify important policymakers
- Identify the privacy officer
- Identify privacy and security officers
- Identify relevant scientific personnel
- Identify relevant staff and students
- Identify students to involve in learning analytics development
- Involve the University Student Board in learning analytics development
- Involve the University Student Board and Works Council in learning analytics development
- Know where to get people with the right skills and knowledge

### People – Training

- Let learning analytics staff write instructional manuals.
- Regularly update instructional manuals and video’s
- Construct manuals in such way that rewriting the entire manual again is not necessary in case of minor changes.
- Provide training like statistics courses.
- Ensure that learning analytics staff has suitable education to fulfill their jobs

- Ensure that learning analytics staff is well trained.
- Organize courses together with people from, for example, the math faculty.
- Let the learning analytics team decides what skills and knowledge need further development.
- Ensure that subject matter of statistics courses is easily applicable for learning analytics staff to mitigate the degradation of knowledge over time.
- Get education from members of the math faculty about how to perform certain analyses.
- Let learning analytics staff earn relevant diplomas or certificates.
- Make a training program, have a subscription on online courses, and follow lectures and conferences.
- Realize that learning analytics involve a lot of on-the-job-training.
- Provide training for users (e.g., how to operate the tools, how to interpret data, how to transfer data into action).
- Let the learning analysts walk the users through the analysis and explain everything.
- Train the tutors and academic advisors on how to interpret our reports.
- Provide learning analytics training to our own employees
- Support users in the use of a dashboard

**Privacy & Ethics – Ethics**
- It is important to understand what you want to do instead of what you can do.
- Understand the ethical considerations
- Have policies on ethical use of data and effective systems of data governance in place first
- Anticipate ethical dilemmas, establish a data policy and governance processes
- Consider establishing an ethics committee

**Privacy & Ethics – Human Decision-making**
- Account for the human dimensions of analytics, not only Artificial Intelligence/ Machine Learning models
- Produce reports/displays that are actionable by educators and students
- Only inform and not make decisions from an algorithm without human evaluation (danger of creating a data driven self-fulfilling prophecy education system)

**Privacy & Ethics – Legal Compliance**
- Ensure that data is anonymized
- Use pseudonymization to make data as unrecognizable as possible
- Do not export notes from meetings with academic advisors, as these are confidential.
- Do not use personal data for learning analytics purposes

- Exclude personal data that is irrelevant and ask for consent.
- Split data sets so different users see different parts of the data.
- Before working on a dataset, ask the supplier to remove all unnecessary information.
- Make clear if third parties will get access to the collected data at any point
- We asked consent and explained for what exact reasons data were collected.
- Explain what an algorithm does
- Allow students to delete any optional/personally supplied information
- Evaluate existing legal framework and its applicability for learning analytics
- Ensure that everything needs to be in line with privacy.
- Ensure that students have the right to be forgotten.
- Define what an algorithm may do from a privacy perspective
- Ask for consent

**Privacy & Ethics – Security**
- Let authorization of access to the network drive be managed by one of the learning analysts
- Guarantee the security of the data
- Have information security policies and practices that are sufficiently robust to safeguard the use of data for analytics
- Have policies that specify rights and privileges regarding access to institutional and individual data

**Privacy & Ethics – Transparency**
- Explain what an algorithm does
- Make research reproducible through references datasets, shared models etc.
- Be transparent about the use of data points and applied algorithms
- Evaluate for generality so that the contexts in which they can be used reliably are known and guard against invalid application
- Be clear and critical in the ways in which it conceives of ‘student success’
- Technology – Automation
- Automate a process for continuous student monitoring
- Write a script for everything that is done with data
- Much is automated, also the delivery of new data sets
- Automate processes so when a new data set arrives, the scripts automatically use that new data
- Use an order in which scripts are executed
- Automate as much as possible so people don’t need to spend time on processing data

- Automate the reading, manipulating, and linking of data, and the making of new models
- Conduct automatic checks after each step of data processing
- Automatically check all data sets.
- Use an automatic method of maintaining data consistency.
- Reduce human error by automating processes
- Automatically notify users of critical issues.

**Technology – Connectivity**

- Connect all remote, branch, and mobile offices to the central office for analytics
- Enable sharing of seamless Analytics-driven information across your organization, regardless of the location.
- Utilize open systems network mechanisms to boost analytics connectivity.
- Always be connected with the network drive
- be available as a service that works with the institution's other student systems
- When other institutions use different systems, exchange of data is difficult
- Allow the information to be shared in the cloud-based data warehouse with other institutions

**Technology – Infrastructure**

- Use Learning Management Systems like Canvas as data source
- Use a Learning Management System as data source
- Use Management Information Systems as data source
- Use Enterprise Resource Planning systems like SAP as data source
- Use Management Information Systems and Enterprise Resource Planning systems as data sources
- Build a dashboard in, for example, Tableau
- Use Tableau for visualizations
- Generate detailed reporting in visual ways
- Explore or adopt different data visualization tools
- Store data in data warehouses
- Implement an Enterprise Data Warehouse (EDW). These provides ease of access to data and therefore can facilitate wide access to data analyses and reports
- Store student data into appropriate databases
- Ensure sufficient capacity to store, manage, and analyze increasingly large volumes of data
- Actively deploy new information technologies such as Supplier Relationship Management (SRM)
- Take advantage of new information technologies such as remote conferencing.

- Explore or adopt cloud-based services for processing data and performing analytics
- Explore or adopt new forms of databases such as Not Only SQL (NoSQL) for storing data.
- Explore or adopt parallel computing approaches (e.g., Hadoop) to big data processing
- Develop and utilize quantitative and qualitative analysis tools (e.g., statistical analysis, data mining, text mining and predictive analysis) to facilitate the semi-automated analysis of numerical, semi-structured and unstructured data to discover new actionable insights from patterns in the data
- Develop and utilize quantitative and qualitative analysis tools (e.g., statistical analysis, data mining, text mining and predictive analysis) to facilitate the semi-automated analysis of numerical, semi-structured and unstructured data to extrapolate patterns found in the data to predict what is likely to occur in the future. These tools are particularly useful for addressing less structured problems
- Use programming languages like R
- Use R and Tableau
- Recognize the distributed architecture of interactions and not attempt to force things into a single database
- Have a shared network drive
- Develop the ability to design, deploy and reconfigure the enterprise data infrastructure to maximize data value
- Develop the ability to manage human, technological and process impacts across the organization arising from learning analytics initiatives. This involves managing changes to the systems environment
- Consider providing a safe environment (e.g., a sandbox) for testing or research purposes.
- Evaluate technological infrastructure
- Have the right IT infrastructure
- Ensure that analytics personnel can create very capable decision support systems driven by analytics.
- Review and plan technology infrastructure to support data generation, extraction, warehousing and integration
- Build a sound technical infrastructure
- Use Git for version control
- Utilize Object-oriented technologies to minimize the development time for new analytics applications.
- Use reusable software modules in new analytics model development
- Have the right tools and applications (IT intensity and ease of data capture, plus data availability)
- Explore and/or adopt open-source software for learning analytics

- Ensure to have the right tools and software for analytics
- Use project management software like Jira, Confluence, and Slack
- Provide a structure that enables users to access data to improve decision making
- Create the ability to develop and utilize self-service analysis applications e.g., reports, dashboards, scorecards, online analytical processing (OLAP) and data visualization technologies, which display output in a user-friendly format that is readily understood by non-technical users. These applications are particularly useful for addressing structured problems and facilitate the visual manipulation and exploration of data
- Let end-users utilize object-oriented tools to create their own analytics applications.
- Evaluate resources available for primary users to uptake learning analytics (e.g., access to digital devices)
- Provide multiple analytics interfaces or entry points for external end-users.
- Provide transparent access to all platforms and applications within user interfaces.

**Technology – System Characteristics**

- Ensure that the system allows information to be readily accessible
- Ensure that the system makes information easy to access.
- Ensure that the system makes information very accessible.
- Document what happens in the data warehouse
- Ensure that learning analytics systems are simple to understand
- Provide packages that offer a complete solution for people who have insufficient programming skills
- Avoid complexity of the system as this could be detrimental to its usability
- Use visualization software packages like Tableau
- Be able to respond to changes in learning and teaching practices
- Be flexible enough to encompass diverse teaching practices
- Applications can be adapted to meet a variety of needs during analytics tasks.
- Use flexible programming languages like R
- Have systems that allow for analytical activities.
- Have systems that can be adapted to meet a variety of analytics needs.
- Ensure that the system can be flexibly adjusted to new demands or conditions during the analytics process.
- Ensure flexibility so that the system can address needs as they arise during the analytics process.
- Use authorization for accessing network drives
- Offer a meaningful guarantee that it will not share private information.
- Protect information about personal issues.
- Protect information about personal identity.
- Have systems that operates reliable for the analytics.

- Have systems that performs reliable for the analytics.
- Be available on mobile platforms
- Be cloud-based
- Be standards-based and technology platform agnostic
- Avoid dependency on technologies that are likely to be replaced or quickly become obsolete
- Have workstations with fast processors and enough memory
- Ensure that software applications can be easily transported and used across multiple analytics platforms.
- Realize that system integration is important for leveraging value from learning analytics and is facilitated by the flexible design of technology infrastructure and systems architecture. It also introduces a degree of complexity and therefore should be done with care and careful consideration of the need
- Ensure enough memory can be used to manipulate a dataset of a couple of GBs
- You always need to be connected with the network drive
- Develop tools and algorithms that provide real time comprehensive feedback to students
- Reduce runtime of scripts
- Use learning analytics systems in such way they save a lot of time
- Support real-time processing of multiple learning data streams
- Reduce runtime of processing of requests
- Have system that provide information in a timely fashion.
- Eliminate delay in scoring course credits and processing them
- Enable multiple levels of accommodation of sophistication of students and teachers as ‘users’
- Make sure users can trust the system’s output
- Let have members of the learning analytics team have total control over what they do.
- Have modern, visual interfaces.
- Have visual interfaces.
- Support data visualization that enables users to easily interpret results
- Ensure that the collaboration portal enables work to be able to be shared and intellectual property to be spread throughout the organization.

## Appendix 3: survey items chapter 7

Questions relate to the general perception of the Learning Analytics Capability Model, its perceived usefulness, and ease-of-use. The survey uses a Likert response scale of 1 (totally disagree) to 5 (totally agree).

1. All in all, I find the Learning Analytics Capability Model a useful model.

2. I would like to be able to use the Learning Analytics Capability Model in my work.

3. The Learning Analytics Capability Model provides me with more useful insights and feedback than other similar models I tried/used.

4. The Learning Analytics Capability Model enables me to get insights in the capabilities necessary for the successful uptake of learning analytics in my institution.

5. The information provided by the Learning Analytics Capability Model helps me identify what capabilities need to be (further) developed at my institution.

6. The Learning Analytics Capability Model provides relevant information in what way to operationalize learning analytics capabilities.

7. The Learning Analytics Capability Model is easy to understand.

8. The use of the Learning Analytics Capability Model is intuitive enough.

9. The Learning Analytics Capability Model is overburdened with information.

10. (*Open question*) The Learning Analytics Capability Model is (not) helping me implementing learning analytics at scale at my institution because:

HOGESCHOOL UTRECHT

Open Universiteit